PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 7/00, 7/01, 15/00, 15/85, C07H 21/04 — A1

(11) International Publication Number: **WO 94/28114**

(43) International Publication Date: 8 December 1994 (08.12.94)

(21) International Application Number: PCT/US94/06079

(22) International Filing Date: 27 May 1994 (27.05.94)

(30) Priority Data:
08/070,164 28 May 1993 (28.05.93) US

(71) Applicant: AMERICAN CYANAMID COMPANY [US/US]; One Cyanamid Plaza, Wayne, NJ 07470 (US).

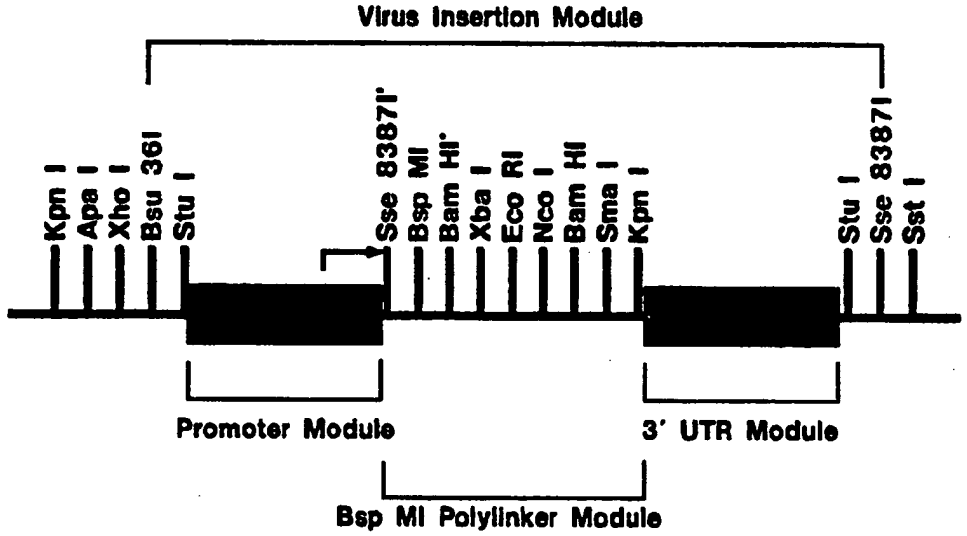
(72) Inventors: WEBB, Nancy, Roberson; 44 Robin Drive, Skillman, NJ 08558 (US). DIERKS, Peter, Michael; 262 Daleview Drive, Yardley, PA 19067 (US).

(74) Agent: GORDON, Alan, M.; American Cyanamid Company, One Cyanamid Plaza, Wayne, NJ 07470 (US).

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: GENE INSERTION BY DIRECT LIGATION *IN VITRO*



(57) Abstract

A method is described for constructing recombinant double stranded DNA viruses, especially baculoviruses and granulosis viruses, by the direct ligation of DNA fragments in vitro. Also described are direct ligation virus vectors, which are insect viruses that have been modified by the insertion of at least one recognition site for a specific restriction endonuclease which does not cut the viral genome, so as to facilitate the insertion of foreign DNA segments by DNA ligation in vitro, and to the recombinant viruses formed by this direct ligation. Further described are modular expression vectors (plasmids) that are designed to facilitate the assembly of gene expression cassettes or other DNA fragments into virus insertion modules (as shown in the figure) which are inserted subsequently into direct ligation virus vectors at predefined sites and in a predefined orientation by ligation in vitro, and to recombinant viruses derived from said modular expression vectors.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

GENE INSERTION BY DIRECT LIGATION IN VITRO

Field Of The Invention

This invention relates to a method for constructing recombinant double stranded DNA insect viruses, especially baculoviruses and granulosis viruses, by the direct ligation of DNA fragments in vitro. This invention further relates to direct ligation virus vectors, which are insect viruses that have been modified by the insertion of at least one recognition site for a specific restriction endonuclease which does not cut the viral genome. This facilitates the insertion of foreign DNA segments by DNA ligation in vitro. This invention also relates to the recombinant viruses formed by this direct ligation. This invention also relates to modular expression vectors (plasmids) that are designed to facilitate the assembly of gene expression cassettes or other DNA fragments into virus insertion modules which are subsequently inserted into direct ligation virus vectors at predefined sites and in a predefined orientation by ligation in vitro, and to recombinant viruses that derive from the use of such modular expression vectors.

Background Of The Invention

The following abbreviations are used throughout this application:

AcMNPV - Autographa californica nuclear polyhedrosis virus

bp - base pairs

BEVS - baculovirus expression vector system
ECV - extracellular virus
GV - granulosis virus
kD - kilodaltons
NPV - nuclear polyhedrosis virus
occ$^-$ - occlusion negative virus(es)
occ$^+$ - occlusion positive virus(es)
OV - occluded virus
PCR - polymerase chain reaction
pfu - plaque forming unit
p.i. - post-infection
PIB - polyhedron inclusion body (also known as occlusion body)
5' UTR: The mRNA or gene sequence corresponding to the region extending from the start site of gene transcription to the last base or basepair that precedes the initiation codon for protein synthesis.
3' UTR: The mRNA or gene sequence corresponding to the region extending from the first base or basepair after the termination codon for protein synthesis to the last gene-encoded base at the 3' terminus of the mRNA.
(+)strand: Refers to the DNA strand of a gene and its flanking sequences which has the same sense as the RNA that is derived from that gene.
(-)strand: Refers to the DNA strand of a gene and its flanking sequences that is complementary to the (+)strand.

Since the advent of recombinant DNA technology, there has been steady growth in the number of systems available for the regulated expression of cloned genes in prokaryotic and eukaryotic cells. One eukaryotic system that has gained particularly widespread use is the baculovirus expression vector

system, or BEVS, developed by Smith and Summers (Bibliography entry 1). This system utilizes a nuclear polyhedrosis virus isolated from the alfalfa looper, Autographa californica, as a vector for the introduction and high level expression of foreign genes in insect cells.

Autographa californica multicapsid nuclear polyhedrosis virus (AcMNPV) is the prototype virus for the Family Baculoviridae. These viruses have large, circular, double-stranded DNA genomes (at least 90-230 kilobases (2)). There are two Subfamilies, Nudibaculovirinae, which do not form occlusion bodies, and the Eubaculovirinae, which are characterized by their ability to form occlusion bodies in the nuclei of infected insect cells. The structural properties of the occlusion bodies are used to further classify the members of this Subfamily into two genera: the nuclear polyhedrosis viruses (NPVs) and the granulosis viruses (GVs).

As exemplified by AcMNPV, the occlusion bodies formed by NPVs are 1-3 microns in diameter and typically contain several hundred virions embedded in a para-crystalline matrix. Occlusion bodies are also referred to as either polyhedra (polyhedron is the singular term) or as polyhedron inclusion bodies (PIBs). The major viral-encoded structural protein of the occlusion bodies is polyhedrin, which has a molecular weight of 29 kilodaltons (kD) (1,3). More than a hundred such occlusions can frequently be found in the nucleus of a single infected cell. GVs are distinguished from NPVs by the fact that their occlusions are much smaller and contain only one virion, which is embedded in a matrix of the viral protein granulin. Nevertheless, the fundamental principles of GV replication are similar to those

described below for AcMNPV.

Viral occlusion bodies play an essential role in the horizontal (insect to insect) transmission of Eubaculovirinae. When a larva infected with AcMNPV dies, large numbers of occlusion bodies are left in the decomposing tissues. In neutral or acidic conditions (pH <10), the protein matrix and outer calyx of the occlusion body protect the embedded virions against chemical degradation in the environment and provide limited protection against UV radiation. However, when the occlusion bodies are ingested by another larva, they dissolve rapidly in the larval midgut, which is strongly alkaline (pH 10.5-12), and the embedded virions are released. These virions then adsorb to and infect various types of midgut cells.

Infected midgut cells synthesize few if any new occlusion bodies. Instead, they produce a second form of the virus, known as extracellular virus (ECV). Whereas the occluded form of the virus is responsible for the horizontal transmission of the virus among larvae, the ECV is used to spread the infection from tissue to tissue internally. This is an essential aspect of normal viral pathogenesis and continues until most tissues of the larva have been infected and lysed. As the virus spreads internally, many of the infected cells, especially hemocytes and fat body cells, produce not only more ECV, but also copious amounts of occluded virus (OV) in the form of occlusion bodies. When the larva dies, the occlusion bodies are deposited in the environment and the cycle begins anew.

Although ECV and OV are genetically identical, they are biochemically distinct. Shortly after the AcMNPV infects a cell, the nucleocapsid

structure (which contains the DNA genome) migrates to the nucleus of the cell, where it is uncoated. This sets in motion a regulated cascade of viral gene expression which leads to the onset of viral DNA synthesis (at about 6 hours post-infection (p.i.)) and the formation of many new nucleocapsids. ECV production begins at about 10-13 hours p.i. with the budding of the nucleocapsids through the cytoplasmic surface of the cell. During the budding process, the nucleocapsids acquire a lipid membrane, or envelope, which is decorated with a viral glycoprotein known as gp64. This protein is specific to the ECV form of the virus and is required for ECV infectivity. The formation of occlusion bodies begins much later (24-36 hours p.i.) and requires the concerted action of numerous specialized viral gene products, the most prominent of which is polyhedrin.

The polyhedrin gene plays a central role in the BEVS technology. Because large amounts of polyhedrin are required for occlusion body formation, the polyhedrin gene is one of the most actively transcribed genes in the viral genome during the very late phases of virus replication. Smith and Summers (1) show that high level expression of a heterologous gene can be achieved by substituting the coding region of the polyhedrin gene with the coding region of a heterologous gene of interest. Since polyhedrin is not required for ECV formation, the resulting virus is able to replicate normally in cultured insect cells. However, it is no longer able to produce polyhedrin for occlusion body formation and is therefore occlusion-negative (occ^-).

The BEVS has been used successfully to express foreign genes isolated from a wide range of prokaryotic and eukaryotic organisms and viruses.

Some representative examples include the human α- and β-interferons, the Drosophila Krueppel gene product, E. coli β-galactosidase, various HIV structural proteins, and a Neurospora crassa site-specific DNA binding protein (3). In general, these genes may encode cytosolic proteins, nuclear proteins, mitochondrial proteins, secreted proteins or membrane-bound proteins. In most cases, the proteins are biologically active and undergo appropriate post-translational modification, including proteolytic processing, glycosylation, phosphorylation, myristylation and palmitylation. Hence, this system has proven to be a highly valued tool for both fundamental molecular research and for the production of proteins for commercial purposes.

Using BEVS technology, recombinant viruses are produced in cultured insect cells by homologous DNA recombination between AcMNPV DNA and a plasmid-based transfer (or transplacement) vector containing the heterologous gene of interest under the control of the polyhedrin gene promoter. To facilitate homologous DNA recombination the modified polyhedrin gene of the transfer vector is flanked at each end by several kilobases (2-4 kb is typical) of native AcMNPV DNA. Many transfer vectors conforming to this general specification have been described (4).

In a typical experiment, purified AcMNPV DNA and transfer vector DNA are mixed together and then transfected into Sf9 insect cells. Once the DNA reaches the cell nucleus, it can be acted upon by cellular proteins involved in the transcription, replication, topological management and repair of DNA. Most of the viral DNA is used without modification as a substrate for viral replication; however, a small fraction (typically 0.1-5%) undergoes homologous

recombination with the transfer vector prior to the onset of virus replication. The product of this recombination event is a virus in which the wild type polyhedrin gene has been transplaced by the desired heterologous gene of the transfer vector. These recombinant viruses can be identified visually with low magnification light microscopy as occ' plaques in a standard viral plaque assay.

An important technical disadvantage of the conventional homologous recombination method is the poor efficiency of recombinant virus production. Recently, two significant modifications to this methodology have been described, which are aimed at increasing the representation of recombinant viruses among the total progeny virus produced in a typical co-transfection experiment.

Kitts et al. (5) show that linearization of AcMNPV viral DNA in the polyhedrin gene region prior to co-transfection with the transfer vector increases the percentage of recombinant viruses to about 30-40% of the total viral progeny. The rationale for this approach is that linearization of the viral DNA provides a better substrate for homologous DNA recombination and disrupts the biological integrity of the viral genome, thereby reducing the background of non-recombinant progeny virus. Site-specific linearization of the viral DNA is achieved by constructing a derivative of AcMNPV that contains a single recognition site for restriction endonuclease Bsu 36I, which does not cut wild type AcMNPV DNA. This site is introduced into the polyhedrin gene region with the aid of a transfer vector in which part of the polyhedrin coding region has been replaced by the E. coli lacZ gene. The lacZ gene encodes the enzyme β-galactosidase and naturally contains a single

Bsu 36I recognition site. Since β-galactosidase can be detected in situ by chromogenic assay methods, a virus containing the desired Bsu 36I site is isolated by plaque purification of occ⁻/β-gal⁺ recombinant viruses.

More recently, Kitts and Possee (6) further refined this approach by constructing an AcMNPV derivative (designated BacPAK6) that contains three Bsu 36I sites, one of which lies in an adjacent AcMNPV gene designated ORF 1629. This site is created by site-directed mutagenesis in such a way that the functional integrity of ORF 1629 is preserved in undigested DNA. Since ORF 1629 is essential for AcMNPV replication, the only way to produce a replication-competent virus from Bsu 36I-digested BacPAK6 DNA is by homologous recombination with a transfer vector that contains at least the missing segment of the ORF 1629 gene. As a result, when Bsu 36I-digested BacPAK6 DNA is cotransfected with a standard polyhedrin-based transfer vector (which typically includes a functional ORF 1629 gene), almost 100% of the viral progeny obtained are recombinants.

Other strategies, based on the use of dominant selectable markers such as the bacterial neomycin resistance gene or the apoptosis-inhibiting viral p35 gene, have also been described for the selective enrichment of recombinant viruses formed by homologous recombination (7). However, regardless of whether this approach or the improvements of Kitts and Possee (6) are used to enrich for recombinant viruses, the current methodology for recombinant virus construction has one inherent limitation: its absolute dependence on the use of transfer vectors.

Transfer vectors, by their very nature, are tied to a specific site of gene insertion in the viral

genome. Hence, a separate transfer vector, and a separate screening strategy and/or selection strategy, must be developed for each site of insertion. At the very least, the reliance on transfer vector-based technology requires at least one additional cloning step for each site of insertion. This limitation is all the more evident as the level of engineering sophistication and the potential uses of recombinant baculoviruses have continued to grow. For example, there is currently much interest in exploring the use of recombinant baculoviruses as targeted delivery systems for genes that can be used to disrupt the normal physiology of specific insect pests in agriculture. The commercial development of such viral insecticides will require the optimization of numerous parameters, including not only the choice of critical regulatory sequences (promoters, enhancers, translational signals and the like) but also the site of foreign gene insertion within the viral genome.

Recently, alternative systems for recombinant baculovirus construction in yeast (8) or by site-specific DNA recombination <u>in vitro</u> (9) have been developed. Both of these systems have the disadvantage of introducing extraneous DNA into the viral genome. Hence, these systems are not well suited to applications, such as viral insecticide development or vaccine development, where it is not desirable to incorporate any extraneous DNA segments (e.g., plasmid vector sequences) into the final recombinant virus.

The current state of the art has given rise to the prevailing opinion that has dominated the conduct of recombinant baculovirus research, which is as follows: "In the past, one of the disadvantages of these [baculovirus expression] systems was the work

involved in isolating a recombinant baculovirus expression vector. The viral genome is too large (130 kb) to manipulate directly, consequently, the standard method for producing virus expression vectors has been to co-transfect insect cells with viral DNA and DNA of a transfer vector modified to incorporate the foreign gene" (6).

Summary Of The Invention

It is an object of the present invention to construct recombinant double stranded DNA insect viruses, such as baculoviruses, by ligation of DNA fragments in vitro.

It is a further object of this invention to provide a means of efficiently inserting a linear DNA fragment into a double stranded DNA insect viral genome without the use of an intermediate transfer vector.

It is an additional object of this invention to provide for the ligation in vitro of a DNA fragment into a double stranded DNA insect viral genome in a predefined orientation.

It is still another object of this invention to provide for a modular expression vector containing a virus insertion module with a nucleic acid sequence encoding a heterologous protein, where said virus insertion module is ligated in vitro into a double stranded DNA insect virus.

It is yet another object of this invention to provide for a modular expression vector which facilitates the ready substitution of promoters, heterologous signals, heterologous genes, and other regulatory sequences in the virus insertion module.

These objects of this invention are achieved

through the construction of a recombinant double stranded DNA insect virus obtained from a double stranded DNA insect virus into which is inserted at least one recognition site for a restriction endonuclease which does not cut the viral genome to create a direct ligation virus vector; and a DNA fragment containing termini selected such that when the direct ligation virus vector is cleaved by the appropriate restriction endonuclease(s), the DNA fragment is ligated in vitro into the direct ligation virus vector. To ligate the DNA fragment into the direct ligation virus vector in a predefined orientation, two different restriction endonuclease recognition sites are inserted into the genome of the double stranded DNA insect virus. It is a preferred embodiment of this invention wherein the DNA fragment is ligated into a region of the insect viral genome which is nonessential for viral replication in cultured cells.

In a particularly preferred embodiment of this invention, the DNA fragment is contained in a modular expression vector. In turn, the modular expression vector comprises a plasmid vector containing a virus insertion module which comprises, in the following order, a recognition site for a restriction endonuclease; a promoter module containing a promoter and a 5' untranslated region (UTR), where the 5' UTR extends from the transcription start site to the last base pair which precedes the translation initiation codon for protein synthesis; a polylinker module to facilitate insertion of a heterologous gene; a 3' UTR module containing at least a site for 3' terminal mRNA processing and polyadenylation; and a recognition site for a restriction endonuclease, such that the two recognition sites permit the ligation in

vitro of the virus insertion module into a direct ligation virus vector in a predefined orientation.

In a further embodiment of the modular expression vector of this invention, the polylinker module of the virus insertion module is altered by the insertion of a nucleic acid sequence encoding a heterologous protein. In a particularly preferred embodiment of this invention, the nucleic acid sequence encoding a heterologous protein is selected from the group consisting of heterologous genes coding for an insect controlling or modifying substance. The modular expression vector may further comprise a nucleic acid sequence encoding a heterologous signal sequence located immediately upstream of the nucleic acid sequence encoding a heterologous protein.

The virus insertion module is ligated in vitro into a direct ligation virus vector. The direct ligation virus vector comprises a double stranded DNA insect virus into which is inserted at least one recognition site for a restriction endonuclease which does not cut the DNA genome of the virus, such that when the direct ligation virus vector is cleaved by the appropriate restriction endonuclease, a DNA fragment (such as the virus insertion module from a modular expression vector) is ligated in vitro into the direct ligation virus vector to produce a recombinant double stranded DNA insect virus.

Brief Description Of The Figures

Figure 1 depicts a linear map of some known baculovirus genes of the AcMNPV genome.

Figure 2 depicts trinucleotide frequencies of a portion of AcMNPV DNA.

Figure 3 depicts tetranucleotide frequencies of a portion of AcMNPV DNA.

Figure 4 depicts the plasmid designated NW33.2, which contains a Bsu-Sse linker inserted into a unique Eco RV site of the AcMNPV Eco RI "I" fragment. The nucleotide sequence of the 40 base pair (bp) linker is also depicted.

Figure 5 depicts the general scheme for assembling the Δp74-1 transfer vector used to assemble a p74-deficient direct ligation virus vector. Figure 5 also depicts the nucleotide sequence of the portion of the transfer vector containing residues 1-69 (beginning at the ATG start codon) of the p74 open reading frame, a 64 bp polylinker (which includes the Bsu-Sse linker) and residues 1287-1937 of the p74 open reading frame (to the TAA termination codon). The leftward and rightward arrows denote the positions of the oligonucleotides used for the PCR-based identification of the p74-deficient direct ligation virus vector.

Figure 6 depicts a portion of a modular expression vector with Bsu 36I and Sse 8387I sites at opposite ends of an expression cassette containing a promoter module, a polylinker module and a 3' UTR module. The polylinker module contains a Bsp MI recognition site. The region bounded by the outermost Bsu 36I and Sse 8387I sites is defined as the virus insertion module.

Figure 7 depicts the scheme for constructing the vector NW46.50, which is a Bsp MI-based modular expression vector containing the AcMNPV 6.9K gene promoter and 3' UTR.

Figure 8 depicts the nucleotide sequences of primers designated NW oligo 1 and PD oligo 23, together with the nucleotide sequence of the fragment

amplified from the AcMNPV Hind III fragment "H" template. Positions are numbered relative to the start site of 6.9K gene translation, which is assigned position +1. Sequences deriving from the synthetic DNA primers are underlined. Nonviral polylinker sequences in the amplified DNA fragment are shown in lower case letters.

Figure 9 depicts the nucleotide sequences of primers designated NW oligo 2 and NW oligo 3, together with the nucleotide sequence of the fragment amplified from the AcMNPV Hind III fragment "H" template. Sequences deriving from the synthetic DNA primers are underlined. The amplified sequences shown in uppercase letters correspond to the complete 3' UTR of the AcMNPV 6.9K gene. Nonviral polylinker sequences in the amplified DNA fragment are shown in lower case letters.

Figure 10 depicts a portion of a modular expression vector with Bsu 36I and Sse 8387I sites at opposite ends of an expression cassette containing a promoter module, a polylinker module and a 3' UTR module. The polylinker module contains an Esp 3I recognition site. The region bounded by the outermost Bsu 36I and Sse 8387I sites is defined as the virus insertion module.

Figure 11 depicts the nucleotide sequences of primers designated DA26FZ and DA26RZ, together with the nucleotide sequence of the fragment amplified from the AcMNPV Pst I fragment "G" template. Positions are numbered relative to the start site of DA26 gene translation, which is assigned position +1. Sequences deriving from the synthetic DNA primers are underlined. Nonviral polylinker sequences in the amplified DNA fragment are shown in lower case letters.

Figure 12 depicts the nucleotide sequences of primers designated 35KPRO1 and 35KPRO2, together with the nucleotide sequence of the fragment amplified from the AcMNPV Hind III fragment "K" template. Positions are numbered relative to the start site of 35K gene translation, which is assigned position +1. Sequences deriving from the synthetic DNA primers are underlined. Nonviral polylinker sequences in the amplified DNA fragment are shown in lower case letters.

Figure 13 depicts the nucleotide sequences of primers designated 69KFZ and 69KRZ, together with the nucleotide sequence of the fragment amplified from the AcMNPV Hind III fragment "H" template. Positions are numbered relative to the start site of 6.9K gene translation, which is assigned position +1. Sequences deriving from the synthetic DNA primers are underlined. Nonviral polylinker sequences in the amplified DNA fragment are shown in lower case letters.

Figure 14 depicts the nucleotide sequences of primers designated PHF and PHR, together with the nucleotide sequence of the fragment amplified from the AcMNPV Hind III fragment "F" template. Positions are numbered relative to the start site of polyhedrin gene translation, which is assigned position +1. Sequences deriving from the synthetic DNA primers are underlined. Nonviral polylinker sequences in the amplified DNA fragment are shown in lower case letters.

Figure 15 depicts the nucleotide sequences of primers designated DMHSP70F and DMHSP70R, together with the nucleotide sequence of the fragment amplified from the Drosophila melanogaster locus 87C1. Positions are numbered relative to the start site of

hsp70 gene translation, which is assigned position +1. The vertical arrow at position -242 marks the major transcription start site of the hsp70 gene. Sequences deriving from the synthetic DNA primers are underlined. Polylinker sequences in the amplified DNA fragment are shown in lower case letters.

Figure 16 depicts the nucleotide sequences of primers designated 35KHR5A and 35KHR5B, together with the nucleotide sequence of the fragment amplified from the AcMNPV Hind III fragment "Q" template. The notation "35K TER" marks the termination codon for the 35K gene. The vertical arrow marks the site of poly(A) addition for the 35K gene. The notation "EcoRI" marks the position of the characteristic IR24 repeat element of hr5. Sequences deriving from the synthetic DNA primers are underlined. Nonviral polylinker sequences in the amplified DNA fragment are shown in lower case letters.

Figure 17A depicts the polymerase chain reaction (PCR) strategy for the amplification of a cuticle signal/codon optimized AaIT gene, which is then digested with Bam HI. Figure 17B depicts the nucleotide sequences of primers designated PD oligo 31 and PVLReverse, which are used to amplify sequences from a plasmid designated pAC0055.1, as well as the nucleotide and amino acid sequences of the cuticle signal and AaIT in the amplified product. Sequences deriving from the synthetic DNA primers are underlined.

Figure 18 depicts a schematic representation and the complete predicted nucleotide sequence of a portion of a modular expression vector (AC0076.1) formed by inserting the cuticle signal/codon optimized AaIT into pMEV1, which contains the AcMNPV DA26 promoter.

Detailed Description Of The Invention

This invention is directed to the construction of a recombinant double stranded DNA insect virus obtained from a double stranded DNA insect virus into which is inserted at least one recognition site for a restriction endonuclease which does not cut the DNA genome of the virus to create a direct ligation virus vector; and a DNA fragment containing termini selected such that when the direct ligation virus vector is cleaved by the appropriate restriction endonuclease(s), the DNA fragment is ligated *in vitro* into the direct ligation virus vector.

The double stranded DNA insect viruses which are modified in accordance with this invention include double stranded enveloped DNA viruses such as (Subfamily, then species) *Entomopoxvirinae* (*Melolontha melolontha* entomopoxvirus), *Eubaculovirinae* (*Autographa californica* MNPV; *Heliocoverpa zea* NPV; *Trichoplusia ni* GV), *Nudibaculovirinae* (*Heliocoverpa zea* NOB), *Ichnovirus* (*Campoletis sonorensis* virus), and *Bracovirus* (*Cotesia melanoscela* virus), as well as double stranded nonenveloped DNA viruses such as the Family *Iridoviridae* (*Chilo* iridescent virus). These insect viruses typically have genomes at least 90 kb in size (2).

Over 400 baculovirus isolates have been described. The Subfamily of double stranded DNA viruses *Eubaculovirinae* includes two genera, nuclear polyhedrosis viruses (NPVs) and granulosis viruses (GVs), which are particularly useful for biological control because they produce occlusion bodies in their life cycle. Examples of NPVs include *Lymantria dispar*

NPV (gypsy moth NPV), Autographa californica MNPV, Anagrapha falcifera NPV (celery looper NPV), Spodoptera littoralis NPV, Spodoptera frugiperda NPV, Heliothis armigera NPV, Mamestra brassicae NPV, Choristoneura fumiferana NPV, Trichoplusia ni NPV, Heliocoverpa zea NPV, Rachiplusia ou NPV, etc. Examples of GVs include Cydia pomonella GV (coddling moth GV), Pieris brassicae GV, Trichoplusia ni GV, Artogeia rapae GV, Plodia interpunctella GV (Indian meal moth), etc. Examples of entomopox viruses include Melolontha melolontha EPV, Amsacta moorei EPV, Locusta migratoria EPV, Melanoplus sanguinipes EPV, Schistocerca gregaria EPV, Aedes aegypti EPV, Chironomus luridus EPV, etc.

The Autographa californica nuclear polyhedrosis virus (AcMNPV) is the prototype virus of the Family Baculoviridae and has a wide host range. The AcMNPV virus was originally isolated from Autographa californica, a lepidopteran noctuid (which in its adult stage is a nocturnal moth), commonly known as the alfalfa looper. AcMNPV has an approximately 130 kb genome (10). This virus infects 12 families and more than 30 species within the order of Lepidopteran insects (11).

Although the invention will be exemplified for Autographa californica NPV (AcMPNV), it is understood that the concepts described herein are applicable for all the above-listed insect viruses. It is further contemplated that the present invention will be highly useful in improving new insect viruses which are not yet identified and classified in the literature.

In the first step in carrying out this invention, a double stranded DNA insect virus is selected and at least one recognition site for a

restriction endonuclease(s) which does not cut the DNA genome of the virus is inserted into the viral genome to create a direct ligation virus vector. The insect virus which will be used to construct the direct ligation virus vector can be either a native virus which lacks the recognition site being inserted or a virus which is first modified to destroy any such preexisting recognition sites.

Subsequently, a DNA fragment containing termini selected such that when the direct ligation virus vector is cleaved by the appropriate restriction endonuclease(s), the DNA fragment is ligated in vitro into the direct ligation virus vector to produce a recombinant double stranded DNA insect virus. As will be discussed below, inserting two different recognition sites into the direct ligation virus vector permits the insertion of a DNA fragment in a predefined orientation.

It is known that there are no recognition sites for the restriction endonuclease Bsu 36I in the genome of the C6 strain of AcMNPV (5). Applicants have also discovered that this site is also not present in the genome of the E2 strain of the virus. Furthermore, Applicants have discovered that there are no recognition sites for the restriction endonuclease Sse 8387I in the genome of AcMNPV strain E2. It is also known that there are no recognition sites for the restriction endonuclease Sma I in the genome of Spodoptera frugiperda NPV (12).

Other recognition sites lacking in these viruses, as well as other viruses, are identified by first analyzing tri- and tetranucleotide sequence frequencies represented in sequenced viral DNA. Codon frequency tables are constructed for each viral genome, such as the tables for portions of AcMNPV

depicted in Figures 2 and 3. The frequency data is used to estimate the likelihood that other enzymes, including those with non-palindromic recognition sequences, might be useful for introducing unique restriction sites into the viral genome. The available viral DNA sequences are then searched for the recognition sequences of the candidate enzymes. If no such sequences are found, the candidate enzyme is tested for its inability to cleave the viral genome. If the enzyme cannot digest the viral genome, it can be used for the construction of a direct ligation virus vector.

Even if sequence information is lacking for a given virus, known sequence information in related viruses is used to identify candidate enzymes. For example, the trinucleotide frequencies listed in Figure 2 for AcMNPV suggest that the recognition site for Sma I (CCCGGG) should occur infrequently in the AcMNPV genome. While there are four such sites in AcMNPV, there is only one such site in Spodoptera littoralis NPV and none in Spodoptera frugiperda NPV.

To ligate the DNA fragment into the direct ligation virus vector in a predefined orientation, two different restriction endonuclease recognition sites are inserted into the double stranded DNA insect viral genome. In a preferred embodiment of this invention, the DNA fragment is inserted at any region of the viral genome of the direct ligation virus vector which is nonessential for viral replication in cultured cells. Examples 1-3 below describe the construction of two different direct ligation AcMNPV viral vectors where the DNA fragment will be inserted in a nonessential region.

In Examples 1 and 2, unique Bsu 36I and Sse 8387I recognition sites are engineered into a viral

construct at the unique Eco RV restriction enzyme recognition site present in the AcMNPV EcoRI "I" fragment. This Eco RV site is located 92 bp upstream of the translation initiation codon of the AcMNPV polyhedrin gene. The direct ligation virus vector is obtained by homologous DNA recombination in cultured insect cells cotransfected with the transfer vector containing the Bsu 36I/Sse 8387I sites and viral DNA modified by substitution of part of the polyhedrin gene so as to be unable to produce occlusion bodies. Because the transfer vector contains a functional polyhedrin gene, homologous recombination yields occlusion-positive viruses which are readily isolated by plaque purification. One such virus is the direct ligation virus vector designated 6.2.1.

In Example 3, unique Bsu 36I and Sse 8387I recognition sites are engineered into the p74 gene region of a transfer vector designated Δp74-1. The direct ligation virus vector designated A4000 is obtained by homologous recombination in cultured insect cells cotransfected with the Δp74-1 transfer vector and wild-type AcMNPV (strain E2) viral DNA. Although A4000 is still able to replicate, this disruption of the integrity of the p74 open reading frame results in the formation of a p74-defective virus which is unable to produce orally infectious occlusion bodies (13).

The direct ligation virus vector is ligated in vitro to a DNA fragment to produce the recombinant double stranded DNA insect virus. The termini of the DNA fragment should be compatible with the termini formed by the digestion of the direct ligation virus vector with the one or two uniquely cutting restriction enzymes. Compatible termini are those which are readily joined by DNA ligase in vitro.

Alternatively, the DNA termini can be modified, for example, by filling in one or more bases to make their ends compatible.

In a particularly preferred embodiment of this invention, the DNA fragment is contained in a modular expression vector. In turn, the modular expression vector comprises a plasmid vector containing a virus insertion module which comprises, in the following order, a recognition site for a restriction endonuclease; a promoter module containing a promoter and a 5' untranslated region (UTR), where the 5' UTR extends from the transcription start site to the last base pair which precedes the translation initiation codon for protein synthesis; a polylinker module to facilitate insertion of a heterologous gene; a 3' UTR module containing at least a site for 3' terminal mRNA processing and polyadenylation; and a recognition site for a restriction endonuclease, such that the two recognition sites permit the ligation in vitro of the virus insertion module into a direct ligation virus vector.

Typically, the virus insertion module is contained in a plasmid vector. For use in direct ligation, the plasmid vector is digested with the unique restriction endonuclease(s) to excise the virus insertion module, which is then inserted by ligation in vitro into a direct ligation virus vector (such as AcMNPV strains 6.2.1 or A4000) previously digested with the same unique restriction endonuclease(s). Examples 4-8 below describe the construction of a variety of modular expression vectors containing a virus insertion module.

Examples 9 and 10 describe the ligation in vitro of a DNA fragment from the modular expression vector NW44.1 into 6.2.1 viral DNA. The two

components are each linearized by digestion with Bsu 36I and Sse 8387I and the small fragments are discarded. The linearized components are then ligated with T4 DNA ligase. The resulting recombinant virus contains the desired insert from the DNA fragment of NW44.1, which indicates that the direct ligation procedure is successful.

Turning to the specific elements of the virus insertion module, the recognition sites at the 5′ and 3′ ends are as described previously. The promoter module contains a promoter and a 5′ untranslated region (5′ UTR), where the 5′ UTR extends from the transcription start site to the last base pair which precedes the translation initiation codon for protein synthesis. For use with AcMNPV, a variety of AcMNPV promoters and 5′ UTR are used, including the "late" 6.9K promoter, the "early" DA26 and 35K promoters, and the "very late" polyhedrin promoter. In some cases, it may be advantageous to use an non-viral promoter rather than a viral promoter. An example of such a non-viral promoter is the Drosophila melanogaster hsp70 (major heat shock) gene promoter. Synthetic promoters and chimeric promoters are also used.

It is not necessary that a complete 5′ UTR be included in the promoter module. It is merely required to have a sufficient portion of the 5′ UTR to allow translation of the heterologous gene in the polylinker module.

The polylinker module serves as the framework into which a heterologous gene is inserted. The insertion of a nucleic acid sequence encoding a heterologous protein necessarily alters the polylinker module, but does not alter the structure or function of the promoter module. Any heterologous gene which

contains an open reading frame flanked by in-frame translation initiation and termination codons may be inserted into the polylinker module. Therefore, all the heterologous genes which have been expressed in the BEVS are also expressible in accordance with this invention.

In a particularly preferred embodiment of this invention, the nucleic acid sequence encoding a heterologous protein is selected from the group consisting of heterologous genes coding for an insect controlling or modifying substance. The insect controlling or modifying substance is selected from the group consisting of toxins, neuropeptides and hormones, and enzymes.

Such toxins include a toxin from the mite species Pyemotes tritici (14), the toxin AaIT from Androctonus australis (15), a toxin isolated from spider venom (16), a toxin from Bacillus thuringiensis subsp. aizawai (17), and a toxin from Bacillus thuringiensis CryIVD (18). Such neuropeptides and hormones include eclosion hormone (19), prothoracicotropic hormone, adipokinetic hormone, diuretic hormone and proctolin (20). An example of an enzyme is juvenile hormone esterase (21).

Although the invention will be exemplified for AaIT, it is understood that the concepts described herein are applicable for all the above-listed insect controlling or modifying substances, as well as for other heterologous proteins. The native nucleotide sequence for the gene encoding AaIT (SEQ ID NO:45) may be used. However, a modified nucleotide sequence may also be used.

The degeneracy of the genetic code permits variations of the nucleotide sequence, while still producing a polypeptide having the identical amino

acid sequence as the polypeptide encoded by the native DNA sequence. The procedure known as codon optimization provides one with a means of designing such an altered DNA sequence.

The design of codon optimized genes should take into account a variety of factors, including the frequency of codon usage in an organism, nearest neighbor frequencies, RNA stability, the potential for secondary structure formation, the route of synthesis and the intended future DNA manipulations of that gene. One such codon optimized AaIT nucleotide sequence is that set forth in SEQ ID NO:29, nucleotides 49-258.

In some cases, the heterologous gene inserted into the polylinker module may include a nucleotide sequence encoding a signal peptide. Signal sequences are required for a complex series of post-translational processing steps which result in secretion of a protein. If an intact signal sequence is present, the protein being expressed enters the lumen of the rough endoplasmic reticulum and is then transported through the Golgi apparatus to secretory vesicles and is finally transported out of the cell. Generally, the signal sequence immediately follows the initiation codon and encodes a signal peptide at the amino-terminal end of the protein to be secreted. In most cases, the signal sequence is cleaved off by a specific protease, called a signal peptidase. Signal sequences improve the processing and export efficiency of recombinant protein expression using viral expression vectors. Where the heterologous protein is an insect controlling protein, optimized expression of the insect controlling protein using an appropriate signal sequence achieves more rapid lethality than wild-type insect virus.

If the native AaIT gene is used, it typically will be immediately downstream of the native AaIT signal peptide, which is encoded by the nucleotide sequence of SEQ ID NO:28. However, it is possible to use a heterologous signal peptide, particularly a signal peptide from an insect species. Seven such insect signal peptides are as follows (listed by type, species, codon optimized and native sequence identification numbers): the cuticle signal sequence from Drosophila melanogaster (SEQ ID NOS:29, nucleotides 1-48;40), the chorion signal sequence from Bombyx mori (SEQ ID NOS:38,39), the apolipophorin signal sequence from Manduca sexta (SEQ ID NOS:36,37), the sex specific signal sequence from Bombyx mori (SEQ ID NOS:43,44), the adipokinetic hormone signal sequence from Manduca sexta (SEQ ID NOS:34,35), the pBMHPC-12 signal sequence from Bombyx mori (SEQ ID NOS:32,33) and the esterase-6 signal sequence from Drosophila melanogaster (SEQ ID NOS:41,42).

In Example 11 below, fragments derived from various modular expression vectors containing the gene for AaIT are ligated in vitro into the direct ligation virus vectors 6.2.1 and A4000. Seven of the eight resulting recombinant viruses test positive by PCR for the formation of derivatives of 6.2.1 and A4000, indicating that the direct ligation is successful. Further confirmation is provided by injection and oral feeding bioassays with Heliothis virescens larvae. Greater than 95% of all responding larvae infected with viruses containing the AaIT gene exhibit contractile paralysis prior to death. Moreover, all of the recombinant AaIT viruses have a shorter mean response time (RT_{50}) than wild type AcMNPV (strain E2).

Further engineering of the polylinker module is also provided herein. The insertion of an

additional restriction endonuclease recognition site in the virus insertion module facilitates the "perfect" fusion of the translation start site of the heterologous gene with the 3' terminus of the promoter module. Specifically, the additional site is located an appropriate distance downstream of the 3' terminus of the 5' UTR and the restriction endonuclease is of the type which cuts both strands of DNA at sites which lie outside of its recognition site.

The position and orientation of the recognition site are set, such that when the digestion products are treated with the Klenow fragment of E. coli DNA polymerase I in the presence of dNTPs, one DNA end produced by this process corresponds exactly to the 3' end of the promoter module. This fragment is then joined directly to the 5' end of the heterologous gene without the introduction of extraneous linker sequences between the 3' terminus of the 5' UTR and the translation initiation codon. Two such recognition sites are those for the restriction endonucleases Bsp MI (Figure 6) and Esp 3I (Figure 10).

The 3' UTR module of the virus insertion module contains at least a site for 3' terminal mRNA processing and polyadenylation. This module may also contain other types of regulatory sequences, such as enhancers. Examples of suitable 3' UTRs include the AcMNPV 6.9K 3' UTR and the 3' terminus of the AcMNPV 35K gene with the AcMNPV homologous region 5 (hr5). The region hr5 enhances the transcriptional activity of the "early" 35K gene and may be useful for other genetic constructs (22).

Yet further engineering of the virus insertion module is part of this invention. A Stu I restriction endonuclease recognition site is inserted

between the recognition site at the 5′ end of the virus insertion module and the 5′ end of the promoter module. A second Stu I restriction endonuclease recognition site is inserted between the 3′ end of the 3′ UTR and the recognition site at the 3′ end of the virus insertion module. If the nucleic acid sequence encoding a heterologous protein in the polylinker module does not contain a Stu I restriction endonuclease recognition site, then the orientation of the expression cassette within the virus insertion module can be reversed by digestion with Stu I, followed by religation of the digestion products. Example 4 below describes such dual orientation constructs.

The foregoing discussion has been focused on the insertion of a virus insertion module containing a single heterologous gene into a direct ligation virus vector. However, it is also within the scope of this invention that the virus insertion module can contain two or more expression cassettes, each with its own promoter module, heterologous gene and 3′ UTR. The promoter and 3′ UTR used for each gene can be the same or different, in order to control the level and timing of expression for each heterologous gene. Such virus insertion modules are used to create recombinant viruses which express a multitude of heterologous proteins from a single virus insertion site. Alternatively, each expression cassette is assembled in an independent virus insertion module, each of which is then inserted at a different unique site within a single direct ligation virus vector, such that the recombinant virus expresses a multiplicity of heterologous proteins.

In summary, this invention provides for the direct ligation <u>in vitro</u> of foreign DNA into double

stranded DNA insect viral genomes. No transfer vectors are needed. No extraneous DNA sequences, such as plasmid vectors, need be introduced. Fragments derived from the modular expression vectors are ligated in vitro into the direct ligation virus vectors. The modular expression vectors provide for ready substitution of promoters, heterologous signals, heterologous genes and other regulatory sequences in the virus insertion modules. The frequency of recovery of the recombinant insect viruses is typically in the range 25-100%. This high rate facilitates screening for the desired constructs. Purification is simplified because one plaque purification step is eliminated. Overall, the invention facilitates the assembly and construction of recombinant double stranded DNA insect viruses.

In order that this invention may be better understood, the following examples are set forth. The examples are for the purpose of illustration only and are not to be construed as limiting the scope of the invention.

Examples

Unless otherwise noted, standard molecular biological techniques are utilized according to the protocols described in Sambrook et al. (23). Standard techniques for baculovirus growth and production are utilized according to the protocols described in Summers and Smith (10). All references to "named" AcMNPV restriction fragments are based on the physical maps of the E2 strain of AcMNPV published in Summers and Smith (10). For example, the designation Eco RI "I" refers to the fragment identified as "I" on the linear map of restriction endonuclease fragments

produced by digestion of AcMNPV strain E2 DNA with Eco RI (Figure 1).

Example 1

Construction of Transfer Vector NW33.2

The concept of constructing recombinant baculovirus genomes by the directional ligation of viral and foreign DNA segments in vitro is predicated on (1) the identification of two or more restriction endonucleases that fail to cut the DNA genome of the baculovirus under study, and (2) the insertion of a synthetic linker containing these sites into a defined location in the baculovirus genome by conventional methods. In some cases, it may be possible to use an enzyme which cuts the viral genome infrequently if the recognition site(s) can be destroyed by site-directed mutagenesis. Optimally, the restriction enzymes should not only have different recognition sequences, but should also (upon cleavage of a susceptible DNA molecule) produce termini that are not readily joined to each other by T4 DNA ligase in vitro. These criteria ensure that the inserted DNA fragment will be joined to the viral DNA in a predefined orientation. Viruses which have been genetically modified to incorporate these special design features are referred to as direct ligation virus vectors.

Two restriction endonucleases are known to meet these criteria for the Autographa californica NPV (AcMNPV) strain E2 genome. Kitts et al. (5) show that there are no recognition sites for restriction endonuclease Bsu 36I in the DNA genome of the C6 strain of AcMNPV. This site is also not present in the DNA genome of AcMNPV strain E2.

Bsu 36I recognizes and cleaves (↓) the DNA

sequence CC↓TNAGG. Those skilled in the art will recognize that numerous isoschizomers of Bsu 36I are known, including Eco 81I, Mst II, Cvn I and Sau I, to name a few. As shown in Figure 2, the Bsu 36I recognition sequence contains two of the four least abundant trinucleotide sequences (AGG and CCT) found in a portion of AcMNPV DNA. Moreover, based on this frequency data one would expect that palindromic restriction sites based on the CCT and AGG trinucleotides (e.g., the Bsu 36I site) would be among the least abundant in the AcMNPV genome. A similar analysis of tetranucleotide sequence frequencies (Figure 3) can be used to identify potential non-cutting enzymes among those restriction endonucleases that recognize 8 bp palindromic sequences. Of the five predicted least abundant classes of palindromic 8-mers, only one class (built on the CCTG and CAGG tetranucleotides) contains the recognition site of a known restriction endonuclease. This enzyme, Sse 8387I, recognizes the sequence CCTGCA↓GG and is the second known enzyme which does not cut the genomic DNA of AcMNPV strain E2. Interestingly, the recognition sequence for Sse 8387I also has the same general sequence pattern (CCT[N_x]AGG) as Bsu 36I (where x=1-5). Those skilled in the art will recognize that the frequency data presented in Figures 2 and 3 can be used to estimate the likelihood that other enzymes, including those with non-palindromic recognition sequences, might be useful for introducing unique recognition sites into the viral genome.

Based on the identification of Bsu 36I and Sse 8387I as enzymes that can be used to prepare direct ligation virus vectors for the E2 strain of AcMNPV, two complementary oligonucleotides (oligo 32 and oligo 33) containing the recognition sites for

these enzymes are chemically synthesized and incorporated into transfer vectors that can be used to insert the sites by a conventional method into predefined locations in the AcMNPV genome. The sequences of these oligonucleotides are:

Oligo 32: 5'-CCTCAGGGCAGCTTAAGGCAGCGGACCGGCAGCCTGCAGG-3' (SEQ ID NO:1)

Oligo 33: 5'-CCTGCAGGCTGCCGGTCCGCTGCCTTAAGCTGCCCTGAGG-3' (SEQ ID NO:2)

In their double-stranded (annealed) configuration, these two oligonucleotides constitute the "Bsu-Sse linker". To anneal the linker, 100 pmol each of oligonucleotides 32 and 33 are diluted into 20 μl of a buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM NaCl and 1 mM EDTA. The mixture is then heated to 78°C for 10 minutes, transferred to 65°C and incubated for 20 minutes, and then slow cooled to room temperature.

One site in which the Bsu-Sse linker is inserted is the Eco RV restriction enzyme recognition site (GAT↓ATC) located 92 bp upstream of the translation initiation codon of the AcMNPV polyhedrin gene. This site is frequently used as a site for foreign gene insertion to produce recombinant viruses that are genetically stable and capable of producing large quantities of recombinant protein (24,25). To construct a plasmid in which the Eco RV site of interest is the only Eco RV site present, the approximately 7 kb AcMNPV Eco RI "I" fragment, which contains the polyhedrin gene, is inserted into the unique Eco RI site of pUC19. One of the resulting clones identified as containing the Eco RI "I" fragment is designated NW32.3. To insert the Bsu-Sse

linker into the unique Eco RV site of NW32.3, 0.3 pmol of NW32.3 DNA are linearized by digestion with Eco RV and ligated to five pmol of the non-phosphorylated double-stranded linker in a volume of 10 microliters. Excess linker is removed by electrophoresing the ligation mixture on a 1% low melt agarose (BioRad, Richmond, CA) gel and isolating the 10 kb DNA band. Approximately one-tenth of this DNA is used to transform competent E. coli strain HB101 cells. One of the resulting plasmids, designated NW33.2, is sequenced to confirm the integrity and orientation of the Bsu-Sse linker (Figure 4).

Example 2
Construction of the Direct Ligation Virus Vector 6.2.1

A recombinant virus containing the Bsu-Sse linker at the Eco RV site upstream of the polyhedrin gene is produced by homologous DNA recombination in cultured Sf9 cells co-transfected with the transfer vector NW33.2 and VL941-500β-gal viral DNA. VL941-500β-gal is a derivative of the E2 strain of AcMNPV in which a part of the polyhedrin gene has been substituted with a segment of DNA that contains the E. coli β-galactosidase gene (26). As such, VL941-500β-gal is unable to produce occlusion bodies (polyhedra) and is phenotypically occlusion-negative (occ^-). Since the virus formed by homologous DNA recombination between NW33.2, which contains a functional polyhedrin gene, and VL941-500β-gal is expected to be occlusion-positive (occ^+), this trait is used to identify potential recombinants.

VL941-500β-gal viral DNA is prepared from

extracellular virus obtained from Sf9 cells infected with VL941-500β-gal virus. Two micrograms of NW33.2 plasmid DNA and 1 μg of VL941-500β-gal viral DNA are co-transfected into Sf9 cells by the calcium phosphate co-precipitation method. Cell supernatants are collected 5 days after transfection and occ⁺ viruses are isolated by three rounds of plaque purification. One occ⁺ plaque, designated 6.2.1, is used to infect 1 x 10^6 Sf9 cells to produce a Passage 1 (P1) virus stock.

The presence of the Bsu-Sse linker in the 6.2.1 viral DNA is verified by PCR analysis of extracellular virus particles using as primers oligo 32 (see Example 1) and "PVLReverse", which anneals to the viral DNA approximately 320 bp downstream of the site of insertion of the Bsu-Sse linker.

PVLReverse: 5'-GGATTTCCTTGAAGAGAGTGAG-3'
(SEQ ID NO:3)

Virus is prepared for PCR analysis essentially as described by Malitschek and Schartl (27). Four microliters of the P1 virus stock is first digested for one hour at 55°C with 200 μg/ml pronase in a 25 μl reaction containing 1X Buffer A (10 mM Tris (pH 8.3), 50 mM KCl, 0.1 mg/ml gelatin, 0.45% Nonidet™ P40 (Shell Oil Co.), and 0.45% Tween™ 20 (ICI Americas)). The pronase is then inactivated by heating to 95°C for 12 minutes. For PCR the pronase-treated virus is mixed with 50 pmol of each of the two oligonucleotide primers in a 50 μl reaction containing 200 μM dNTPs (this is a mixture of nucleotides that contains 200 μM each of dATP, dGTP, dCTP and dTTP), 1.5 mM $MgCl_2$, 1X Buffer A and 2.5 units AmpliTaq™ DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT). The

sample is subjected to 25 cycles of amplification, each consisting of 1 minute at 94°C (denaturation step), 1.5 minutes at 55°C (annealing step), and 2.5 minutes at 72°C (extension step). The 25 cycles are followed by a 7 minutes extension step at 72°C. One-fifth of the reaction mix is electrophoresed on a 1.8% agarose gel to confirm the presence of the 320 bp amplification product.

Viral DNA is isolated from the 6.2.1 extracellular virus and further characterized by restriction enzyme analysis. Digestion with Eco RI, Bsu 36I, and Sse 8387I is used to confirm the presence of the unique Bsu 36I and Sse 8387I sites in the Eco RI "I" fragment of 6.2.1.

Example 3
Construction of p74-Deficient Direct Ligation Virus Vector A4000

A direct ligation virus vector incorporating the Bsu-Sse linker (Example 1) into the p74 gene of AcMNPV is produced by conventional methods using a transfer vector which contains sufficient viral DNA sequences to ensure efficient homologous recombination with wild type AcMNPV DNA. Unlike the 6.2.1 direct ligation virus vector, in which the Bsu-Sse linker is incorporated into an untranslated region of the genome, the Bsu-Sse linker in the current example is inserted into the viral genome in a manner which is intentionally designed to disrupt the integrity of the p74 open reading frame. This results in the formation of a p74-defective virus which produces occlusion bodies that are not orally infectious (13).

Figure 5 shows the general scheme for assembling the transfer vector used to construct the

p74-deficient direct ligation virus vector. The approximately 9.9 kb Bst EII "E" fragment, which contains the p74 gene, is isolated from AcMNPV viral DNA and inserted into a plasmid vector. An approximately 1200 bp Xho I/Hpa I fragment encoding amino acids 24-429 of the p74 protein is replaced with a polylinker containing the Bsu-Sse linker to produce the Δp74-1 transfer vector. The transfer vector Δp74-1 is comprised of 4750 bp of 5' flanking sequences, residues +1 - +69 of the p74 open reading frame, a 64 bp polylinker (which includes the 40 bp Bsu-Sse linker sequence), p74 gene sequences from bp +1287 to the termination codon at +1937 (see Figure 5; SEQ ID NO:4), followed by 1796 bp of 3' flanking sequences. Samples of an E. coli strain HB101 harboring this transfer vector have been deposited by applicants' assignee with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., and have been assigned ATCC accession number 68988.

A recombinant virus containing the Bsu-Sse linker inserted into a mutated p74 open reading frame is produced by homologous recombination in cultured Sf9 cells co-transfected with Δp74-1 and wild type AcMNPV (strain E2) viral DNA. The desired recombinant virus is identified by PCR using one oligonucleotide that is specific for the polylinker sequence in the deleted gene (location denoted by leftward arrow in Figure 5) and a second oligonucleotide that is specific for the 3' end of the p74 gene coding region (location denoted by rightward arrow in Figure 5). Samples of this virus, which is designated AcMNPV strain A4000, have been deposited by applicants' assignee with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852,

U.S.A., and have been assigned ATCC accession number VR 2373.

Example 4

Construction of Bsp MI-based Modular Expression Vectors NW44.1 and NW46.50

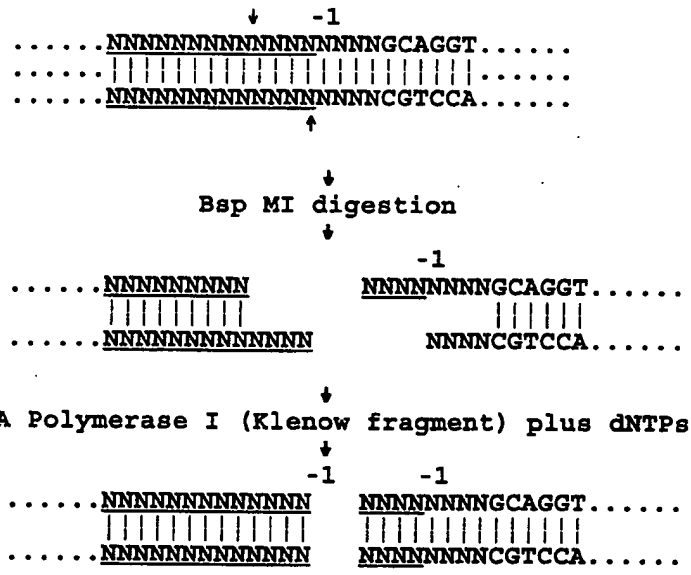
To utilize the special design features of direct ligation virus vectors, the termini of the DNA fragment to be inserted into the viral genome should be compatible (i.e., readily ligated by T4 DNA ligase) with the termini formed by double digestion of the direct ligation vector with the two uniquely cutting restriction enzymes. Figure 6 displays an example of an expression vector design that is intended for use with a direct ligation virus vector such as AcMNPV strains 6.2.1 or A4000. In this example, the Bsu 36I and Sse 8387I recognition sites flank the ends of a tripartite expression cassette that is composed of the following modules: (1) a promoter module, which is used to regulate gene transcription; (2) a polylinker module, which facilitates insertion of the heterologous DNA sequences whose expression is desired; and (3) a 3' untranslated region (3' UTR), which provides a site for primary transcript processing and polyadenylation. The region bounded by the outermost Bsu 36I and Sse 8387I sites is defined as the virus insertion module. The internal Sse 8387I site marked with an asterisk in the polylinker module in Figure 6 is destroyed when the Bsp MI site is used to insert heterologous gene sequences into the modular expression vector. This internal site is eliminated in the pMEV series of vectors described in Example 5. Similarly, the internal Bam HI site marked with an asterisk in Figure 6 is not required and is also

eliminated in the pMEV series of vectors.

Following insertion of the desired heterologous gene sequences into the polylinker module, the virus insertion module is excised from the plasmid vector by double digestion with Bsu 36I and Sse 8387I and inserted by DNA ligation in vitro into a Bsu 36I/Sse 8387I double cut direct ligation virus vector, such as AcMNPV strains 6.2.1 or A4000.

The design depicted in Figure 6 has two additional features worth noting. The first is the presence of Stu I recognition sites at both ends of the tripartite expression cassette. If the heterologous gene sequences inserted into the polylinker module do not contain Stu I sites, the orientation of the entire expression cassette can be reversed in the virus insertion module by digesting the plasmid with Stu I and religating the pieces. The second feature is designed to facilitate the fusion of an exogenous open reading frame (beginning with a suitable translation initiation codon, such as ATG) with the 3′ terminus of a natural or synthetic 5′ untranslated region (5′ UTR), such that no extraneous linker sequences are introduced between the 3′ terminus of the 5′ UTR and the initiation codon. This is accomplished by the precise placement of a Bsp MI recognition site near the 5′ terminus of the polylinker module. Bsp MI belongs to a class of Type II restriction endonucleases that cuts both strands of the DNA duplex at sites which lie outside (and on the same side) of its recognition sequence. Moreover, the cuts in each strand are staggered in such a way that the ends of the fragments have 5′ protruding termini that can be used as template:primer complexes for a DNA polymerase, such as the Klenow fragment of E. coli

DNA polymerase I (23)(SEQ ID NO:5):



As can be seen from this example, if the Bsp MI site is placed in the correct orientation 4 bp downstream of the 3' terminus of the 5' UTR (underlined nucleotides shown above), one DNA end produced by this process corresponds exactly to the 3' end of the 5' UTR. The 5' UTR is then joined directly to a blunt ended-fragment whose sequence begins with the ATG (or other) initiation codon of the desired open reading frame. As described below, such fragments are easily prepared by PCR techniques. By convention, the 3' terminus of the 5' UTR sequence is designated as position -1 and is used as the landmark for all numerical references in the virus insertion module. In all of the vectors described herein, the 5' UTR is incorporated as part of the promoter module and has the complete nucleotide sequence of the 5' UTR naturally associated with the promoter in that module.

The scheme for constructing Bsp MI-based modular expression vectors containing the AcMNPV 6.9K

gene promoter and 3' UTR is illustrated in Figure 7. The 6.9K gene is a "late" gene that encodes a small arginine-rich DNA-binding protein used for packaging viral DNA into nucleocapsids (28,29). The 5' and 3' non-coding sequences flanking the 6.9K open reading frame are isolated by PCR amplification using the AcMNPV Hind III "H" fragment as template. The oligonucleotide primers used for the amplification reactions are each composed of two functional regions. The 5' portion of each oligonucleotide is not homologous to the AcMNPV template and is used to incorporate specific restriction sites into the final PCR product. The 3' portion of each oligonucleotide is homologous to sequences in the AcMNPV genome that define one of the termini of the amplified region.

The oligonucleotides used for PCR amplification of the module containing the 6.9K gene promoter and 5' UTR are NW oligo 1 (SEQ ID NO:6), which anneals to the (-)strand approximately 200 bp upstream of the 6.9K translation initiation codon and contains Xho I, Bsu 36I, and Stu I recognition sites, and PD oligo 23 (SEQ ID NO:7), which anneals to the (+)strand immediately upstream of the 6.9K translation initiation codon and contains Bam HI and Bsp MI recognition sites. The sequences of these primers and of the PCR amplification product (SEQ ID NO:8) are presented in Figure 8.

The oligonucleotides used for PCR amplification of the 6.9K gene 3' UTR are NW oligo 2 (SEQ ID NO:9), which anneals to the (-)strand immediately downstream from the translation stop codon of the 6.9K open reading frame and contains recognition sites for Xba I, Eco RI, Nco I, Bam HI, Sma I and Kpn I, and NW oligo 3 (SEQ ID NO:10), which anneals to the (+)strand approximately 200 bp

downstream of the translation stop codon and contains recognition sites for Stu I, Sse 8387I and Sst I. The sequences of these primers and of the PCR amplification product (SEQ ID NO:11) are presented in Figure 9.

The amplification reactions for the 6.9 K promoter module and 3' UTR module are conducted in separate GeneAmp tubes (Perkin-Elmer Cetus, Norwalk, CT) according to the following procedure. Fifty picomoles of the appropriate primer pair (i.e., 50 pmol of each oligonucleotide) are combined with 250 pg of AcMNPV Hind III fragment "H" DNA (see Figure 1) in a 50 μl reaction mixture containing 200 μM dNTPs, 1.5 mM $MgCl_2$, 1X Buffer A and 2.5 units AmpliTaq™ DNA polymerase (Perkin-Elmer Cetus, Norwalk. CT). The samples are first subjected to 5 rounds of amplification consisting of 1 minute at 94°C (denaturation step), 1.5 minutes at 45°C (annealing step), and 2.5 minutes at 72°C (extension step). This is followed by 20 cycles consisting of 1 minute at 94°C (denaturation step), 1.5 minutes at 60°C (annealing step), and 2.5 minutes at 72°C (extension step). The last extension step is extended an additional 7 minutes.

The amplification products are extracted once with chloroform, once with phenol:chloroform, and then precipitated with ethanol. The fragment containing the presumptive promoter module is digested with Xho I and Bam HI (see Figure 8) and the fragment containing the presumptive 3' UTR module is digested with Xba I and Sst I (see Figure 9). Each of the desired fragments is then isolated by electrophoresis on a 1.8% low melt agarose gel. The promoter fragment is inserted into the polylinker of Bluescript SK+ (Stratagene, La Jolla, CA) between the Xho I and Bam

HI sites, and the 3' UTR fragment is inserted into the polylinker of a separate Bluescript SK+ plasmid between the Xba I and Sst I sites (see Figure 7). The plasmid identified as containing the 6.9K promoter module is designated NW39.2, while that containing the 6.9K 3' UTR is designated NW41.5. Both NW39.2 and NW41.5 are sequenced to verify the integrity of the 6.9K gene segments and flanking linker sequences.

To construct a complete Bsp MI-based modular expression vector, NW39.2 and NW41.5 are digested with Xba I and Sst I and the fragments are resolved by electrophoresis on a 1.2% low melt agarose gel. The 3.1 kb fragment derived from NW39.2 and the 200 bp fragment derived from NW41.5 are extracted from the gel and ligated together. A clone containing the desired promoter, polylinker and 3' UTR modules is identified by restriction enzyme analysis and is designated NW44.1.

To obtain a vector in which the expression cassette has the opposite orientation within the virus insertion module, NW44.1 is first digested with Stu I. The 2.9 kb Stu I fragment is purified by gel electrophoresis, dephosphorylated to prevent self-ligation, and re-ligated to the 450 bp NW44.1 Stu I fragment containing the expression cassette. A clone containing the Stu I insert in the opposite orientation relative to NW44.1 is identified by restriction enzyme analysis and designated NW46.50.

Example 5

Construction of Esp 3I-based Modular Expression Vectors pMEV1, pMEV2, pMEV3 and pMEV4

The modular expression vectors pMEV1, pMEV2, pMEV3 and pMEV4 are constructed from NW46.50 by

substituting the promoter-containing Pst I/Xba I fragment of NW46.50 with Pst I/Xba I-digested fragments containing the AcMNPV DA26 (pMEV1), 6.9K (pMEV2), polyhedrin (pMEV3) and 35K (pMEV4) viral gene promoters. The DA26 and 35K genes are expressed at an early stage in the life cycle of the virus (i.e., before the onset of DNA synthesis) (30,31). As noted earlier, the 6.9K gene encodes a "late" class structural protein, which is expressed after the onset of DNA synthesis (28). The polyhedrin gene belongs to the class of genes that are expressed "very late" in the virus life cycle and encodes the major structural component of the viral occlusion bodies.

As depicted in Figure 10, the design of the Esp 3I-based vectors is a refinement of the Bsp MI-based model, in which (1) the redundant Bam HI and Sse 8387I sites in the polylinker module are eliminated, and (2) the Bsp MI recognition site is replaced by an Esp 3I site. Esp 3I belongs to the same general class of Type II restriction endonuclease as Bsp MI, in that it cuts outside of its recognition sequence and produces 5′ protruding termini that can be filled in by DNA polymerase. Experience indicates, however, that Esp 3I has a more robust activity than Bsp MI and is the preferred enzyme when all other factors permit its use (e.g., when there are no Esp 3I sites in either the promoter module or the 3′ UTR module). To use Esp 3I in the manner illustrated earlier for Bsp MI, its recognition site must be placed in the correct orientation 1 bp downstream of the 3′ end of the 5′ UTR.

As described in Example 4 for the Bsp MI-based vectors, the promoter fragments used in constructing the Esp 3I-based vectors are formed by PCR amplification of cloned viral DNA using promoter-

specific pairs of oligonucleotide primers. The primers are designed so that the amplified promoter segments have the following general structure: (1) a 5′ terminal 22 bp heteropolymeric synthetic sequence with recognition sites for restriction endonucleases Sst I, Sse 8387I and Stu I (in that order); (2) a segment of viral DNA that extends from a point 100-350 bp upstream of the predominant transcriptional start site of the gene to the 3′ terminus of the 5′ UTR (i.e., position -1 with respect to the translation initiation codon); and (3) a 3′ terminal 23 bp heteropolymeric region with recognition sites for restriction endonucleases Esp 3I and Xba I (in that order). The location and orientation of the Esp 3I recognition site places the cleavage sites between positions -5 and -4 in the (+)strand and between positions -1 and +1 in the (-)strand.

The template used to prepare each promoter, the sequences of the primers and the sequences of the amplified PCR products are shown in Figures 11 (DA26 promoter module)(SEQ ID NOS:12-14), 12 (35K gene promoter module)(SEQ ID NOS:15-17), 13 (6.9K gene promoter module)(SEQ ID NOS:18-20) and 14 (polyhedrin gene promoter module)(SEQ ID NOS:21-23).

For each amplification reaction, 50 pmol of the appropriate primer pair are mixed with 250 pg of template DNA in a 50 μl reaction mixture containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM dNTPs, 100 μg/ml gelatin and 2.5 units AmpliTaq™ DNA polymerase (Perkin-Elmer Cetus, Norwalk. CT). For the DA26, 6.9K and polyhedrin promoter modules, the samples are then subjected to 2 rounds of amplification consisting of 1 minute at 94°C (denaturation step), 1.5 minutes at 40°C (annealing step), and 2.5 minutes at 72°C (extension step). This

is followed by 15 cycles of 1 minute at 94°C (denaturation step), 1.5 minutes at 60°C (annealing step), and 2.5 minutes at 72°C (extension step). The last extension step is programmed to run an additional 7 minutes. For the 35K promoter module, the sample is amplified through 25 cycles of 1 minute at 94°C (denaturation step), 1.5 minutes at 55°C (annealing step), and 3.0 minutes at 72°C (extension step). As in other reactions, 7 minutes is added to the last extension step.

Each reaction is terminated by the addition of EDTA to 10 mM and Sarkosyl (sodium N-lauroylsarcosine) to 0.2% (w/v). The products are then extracted once with chloroform, once with phenol:chloroform and precipitated with ethanol. The DNA samples are redissolved in an appropriate buffer and then digested with Pst I (which recognizes the central six basepairs [CTGCA↓G] of the Sse 8387I site) and Xba I. Each presumptive promoter fragment is then purified by gel electrophoresis on a 1.2% low melt agarose gel and ligated to a 3.2 kb Pst I/Xba I vector fragment prepared from NW46.50 (see Figure 7). This fragment contains the polylinker module, 3' UTR module and Bluescript SK+ framework of NW46.50. The desired recombinants are identified by restriction enzyme analysis and DNA sequence determination. Representative isolates of each expression vector are designated pMEV1.1 for the DA26 promoter, pMEV2.1 for the 6.9K gene promoter, pMEV3.1 for the polyhedrin gene promoter and pMEV4.1 for the 35K gene promoter. Samples of an E. coli strain DH5α harboring plasmid pMEV1.1 (AC0064.1) have been deposited by applicants with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., on April 7, 1993, and have been assigned ATCC accession

number 69275.

Example 6

Construction of Esp 3I-based Modular Expression Vectors Containing the Drosophila hsp70 Gene

In some cases, it may be advantageous to use an insect cell promoter rather than a viral promoter to direct transcription of a foreign gene in a baculovirus-based expression system. In particular, Morris and Miller (32) report that a Drosophila hsp70 (major heat shock) gene promoter functions at a comparable or better level than the AcMNPV ETL (early) gene promoter in directing expression of a chloramphenicol acetyltransferase (CAT) reporter gene in a variety of insect cell lines. To construct an expression vector that uses an insect cell promoter to control foreign gene expression and that can be used with the direct ligation virus vectors, a DNA segment of Drosophila melanogaster DNA containing the hsp70 promoter and 5' UTR is amplified by PCR and substituted into one of the modular expression vectors described above. The sequences of the primers for this reaction and the predicted sequence of the amplified fragment, which contains an Esp 3I site in the presumptive polylinker region, are presented in Figure 15 (SEQ ID NO:24). The procedures for PCR amplification of the hsp70 promoter module, and for inserting this module into an NW46.50-based expression vector are as described in Example 5 for the AcMNPV 35K promoter module. A resulting clone with the desired structure is designated pMEV5.

Example 7
Construction of Esp 3I-based Modular Expression Vectors Containing an Alternative 3' UTR and the hr5 enhancer

AcMNPV contains five regions of homologous DNA sequence, designated hr1 to hr5, that are widely interspersed along its genome (22,33). Each region is 500-800 bp in length and contains variations of several repeated sequence motifs, one of which (IR24) (34) contains an Eco RI recognition site. Functional studies have shown that regions such as hr5 are complex cis-acting regulatory domains that can enhance the transcriptional activity of at least some linked early viral genes (e.g., the 35K gene) by as much as 300- to 1000-fold (35). In addition, recent evidence suggests that the hr elements may also serve as origins for DNA synthesis (36).

The hr5 element lies downstream and immediately adjacent to the AcMNPV 35K gene, and is therefore well suited for use as an alternative 3' UTR module. Figure 16 displays the sequences of two oligonucleotides (SEQ ID NOS:25,26) for the PCR amplification of a segment of the AcMNPV genome that begins just upstream of the 3' terminus of the 35K gene and extends through all six IR24 repeats (marked by the Eco RI sites) of hr5. The conditions used for PCR amplification of the hr5 domain are the same as those described for the amplification of the 35K promoter module in Example 5. After purification, the PCR product (SEQ ID NO:27) is digested with Bam HI and Xho I and the presumptive hr5 enhancer module is isolated on a 1% low melt agarose gel. This module is then ligated with gel purified Bam HI/Xho I vector fragments prepared from each of the Esp 3I-based

modular expression vectors described in Examples 5 and 6. The result is a new series of vectors in which the 6.9K gene-derived 3' UTR module is replaced by the hr5 module. Plasmids with the desired structures are identified by restriction enzyme analysis and are designated pMEV1A (with the DA26 gene promoter module), pMEV2A (with the 6.9K gene promoter module), pMEV3A (with the polyhedrin gene promoter module), pMEV4A (with the 35K gene promoter module) and pMEV5A (with the Drosophila melanogaster hsp70 promoter module).

Example 8
Construction of Modular Expression Vectors Containing Codon-Optimized or Native Sequence AaIT Genes

One application in which the direct ligation technology and the vectors described in Examples 4-7 are particularly useful is the design and optimization of recombinant viral insecticides. These are viruses, especially baculoviruses, whose insecticidal properties have been enhanced by the addition of one or more foreign genes that encode insect-specific toxins (14,15,37); peptide hormones (38) or enzymes (21).

The peptide AaIT, which is found in the venom of the North African scorpion Androctonus australis, is an example of such an insect-specific toxin (15). When AaIT is injected into the body cavity of an insect larva, it binds selectively to voltage-sensitive sodium channels and causes a contractile paralysis. Chronic administration of the toxin, which can be achieved by infecting insect larvae with AaIT-producing viruses, is associated with

a prolonged state of paralysis and eventual death.

Transfer vectors are constructed for the insertion of AaIT genes into the polyhedrin gene region of AcMNPV. These transfer vectors are derivatives of the published pVL1393 (4) and pVL985 (26) vectors. In one vector, the AaIT coding region is inserted between the Bam HI and Eco RI sites of pVL1393 and has the same nucleotide sequence (SEQ ID NO:45) as the coding region of the AaHIT1 cDNA described by Bougis et al. (39), and is used in conjunction with the native nucleotide sequence encoding the native AaIT signal peptide (SEQ ID NO:28). In seven other vectors, the AaIT gene is inserted into the Bam HI site of pVL985 and consists of a codon-optimized nucleotide sequence (SEQ ID NO:29, nucleotides 49-258) for the mature AaIT toxin, which is linked in each vector to a separate codon-optimized sequence for one of seven different insect signal peptides identified below (SEQ ID NOS:29 (nucleotides 1-48),32,34,36,38,41,43). One such construct, containing the codon-optimized AaIT gene linked to the signal peptide of a Drosophila cuticle gene, is designated pAC0055.1. Figure 17A depicts the region of pAC0055.1 which contains the cuticle signal and AaIT gene inserted into the unique Bam HI site located between residues +34 and +177 of the polyhedrin gene. The ATT sequence at +1 indicates the mutated translation start codon in the parental pVL985 vector. Samples of an E. coli strain HB101 harboring the transfer vector pAC0055.1 have been deposited by applicants with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., and have been assigned ATCC accession number 69166. In an alternative embodiment, the native, rather than codon optimized, nucleotide sequences

encoding the seven different insect signal peptides are used (SEQ ID NOS:33,35,37,39,40,42,44).

The toxin coding segments of these transfer vectors are recovered by PCR for subsequent insertion into modular expression vectors. The PCR strategy (Figure 17A) and the sequence of the amplified fragment (Figure 17B) are exemplified for the Cuticle/AaIT gene. The (+)strand primer used for each reaction is an oligonucleotide of 25-27 bases whose 5' terminus coincides with the ATG translation initiation codon of the gene to be amplified. The specific sequences of the (+)strand primers used to amplify each AaIT gene are listed below, followed by identification of the insect species signal peptide which is the source of the primer used:

5'-ATG AAC TAC GTC GGG CTG GGC CTC ATC-3' (esterase-6 signal from Drosophila melanogaster; SEQ ID NO:41, nucleotides 1-27)

5'-ATG TAC AAA CTG ACC GTC TTC CTG ATG-3' (adipokinetic hormone signal from Manduca sexta; SEQ ID NO:34, nucleotides 1-27)

5'-ATG TTC AAG TTC GTG ATG ATC TGC GCC-3' (cuticle signal from Drosophila melanogaster; SEQ ID NO:29, nucleotides 1-27)

5'-ATG GCC GCT AAA TTC GTC GTG GTT CTG-3' (apolipophorin signal from Manduca sexta; SEQ ID NO:36, nucleotides 1-27)

5'-ATG AAA CTC CTG GTC GTG TTC GCC ATG-3' (pBMHPC-12 signal from Bombyx mori; SEQ ID NO:32, nucleotides 1-27)

5'-ATG CGC GTC CTG GTG CTG TTG GCC TGC-3' (sex specific signal from Bombyx mori; SEQ ID NO:43, nucleotides 1-27)

5'-ATG TTC ACC TTC GCT ATT CTG CTC TTG-3'

(chorion signal from Bombyx mori; SEQ ID NO:38, nucleotides 1-27, except that nucleotide 25 in the primer is a T instead of a C)

5'-ATG AAA TTT CTC CTA TTG TTT CTC G-3'

(native signal for AaIT; SEQ ID NO:28, nucleotides 1-25)

The first seven primers are used with a codon optimized gene encoding AaIT (SEQ ID NO:29, nucleotides 49-258); the last primer is used with the native gene encoding AaIT (SEQ ID NO:45).

The (-)strand primer in each case is the PVLReverse primer (SEQ ID NO:3)(see Example 2), which hybridizes to a common site located about 35-40 bp downstream of the 3' terminus of the AaIT gene (i.e., in the polyhedrin gene). The conditions for the PCR reaction are essentially as described for the 35K promoter module in Example 5. After purification the reaction products are treated with the Klenow fragment of E. coli DNA polymerase I in the presence of all four dNTPs to ensure that the PCR products have blunt 5' termini. The 3' terminus of each toxin coding fragment is then defined by digesting the PCR products with either Bam HI (for the codon-optimized AaIT genes) or Eco RI (for the native sequence AaIT gene), and the fragments are purified by electrophoresis on a 1.8% low melt agarose gel. Figure 17B depicts the complete nucleotide sequence of the PCR amplified codon optimized cuticle/AaIT coding region (SEQ ID NO:29, nucleotides 1-258).

To prepare the Esp 3I-based modular expression vectors for toxin gene insertion, each vector is digested with Esp 3I and the resulting 5' protruding termini are filled in by the action of E. coli DNA polymerase I (Klenow fragment) in the

presence of all four dNTPs. Part of each preparation is then digested with Bam HI (for insertion of the codon-optimized gene fragments) and part is digested with Eco RI (for the native sequence gene fragment). The vector is separated from the liberated polylinker fragment by electrophoresis on a 1% low melt agarose gel and then ligated in separate reactions to the appropriate AaIT-encoding gene fragments. A schematic representation and the complete nucleotide sequence of the virus insertion module designated AC0076.1 (SEQ ID NO:30) formed by inserting the Cuticle/AaIT coding region into pMEV1 (which contains the AcMNPV DA26 promoter) is presented in Figure 18.

Example 9
Preparation of 6.2.1 and A4000 Viral DNAs for Ligation

To prepare the 6.2.1 and A4000 viral DNAs for gene insertion by ligation in vitro, the DNAs are linearized by sequential digestions with Sse 8387I and Bsu 36I, and then separated from the small Bsu-Sse linker fragment by gel filtration chromatography. In a typical preparation, forty micrograms of 6.2.1 or A4000 viral DNA are digested for 2 hours at 37°C with 100 units of Sse 8387I (Takara Biochemical, Inc., Berkeley, CA) in a 250 μl reaction containing 10 mM Tris pH 7.5, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 50 mM NaCl, and 0.01% BSA. The reaction mixture is then adjusted to 100 mM NaCl and 50 mM Tris HCl, pH 7.9, and the DNA is digested for 2 hours at 37°C with 100 units of Bsu 36I (New England Biolabs, Beverly, MA). The reaction is then terminated by adding SDS to a final concentration of 1%(w/v), NaCl to a final concentration of 0.3 M and EDTA to a concentration of

10 mM. Thereafter, the DNA is chromatographed on a "poly-prep" column (BioRad Laboratories, Richmond, CA) containing a 2 ml bed volume of Sephacryl-300 (Pharmacia, Piscataway, NJ) equilibrated with 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.1% SDS, and 0.3 M NaCl. Twelve 150 μl fractions are collected. Ten microliters of each fraction are analyzed by gel electrophoresis to identify fractions containing the viral DNA. These fractions are pooled, extracted once with phenol:chloroform, and the viral DNA is then precipitated with ethanol. The DNA is resuspended in TE (10 mM Tris-HCl (pH 8.0), 1 mM EDTA) at a concentration of 0.2-1 μg/μl and stored at 4°C. To determine if the viral DNA has been linearized completely, an aliquot is digested with Eco RI and analyzed by gel electrophoresis. Viral DNA exhibiting a 7 kb Eco RI fragment has not been digested completely with Bsu 36I and Sse 8387I and is not used.

Example 10

Insertion Of Foreign DNA Into The Direct Ligation Virus Vector 6.2.1 By Ligation In Vitro

The efficiency of obtaining recombinant viruses by ligating foreign DNA in vitro into the unique Bsu 36I/Sse 8387I cloning site of linearized 6.2.1 viral DNA is demonstrated with the 2.9 kb Bsu 36I/Sse 8387I fragment of NW44.1, which consists mainly of DNA sequences from the Bluescript cloning vector (Figure 7).

This fragment is purified by digesting plasmid NW44.1 sequentially with Sse 8387I and Bsu 36I and then separating the digestion products on a 1% low melt agarose gel containing 40 mM Tris-acetic acid (pH 7.8), 1 mM EDTA and 0.5 μg/ml ethidium bromide. After

electrophoresis, a gel slice containing the 2.9 kb Bsu 36I/Sse 8387I vector fragment is carefully excised. To extract the DNA from the gel, the slice is diluted with 3 volumes of a buffer containing 20 mM Tris-HCl (pH 7.5), 0.4 M sodium acetate and 1 mM EDTA. The mixture is heated to 65°C until the gel slice is melted, then cooled to 37°C and extracted with an equal volume of water-saturated phenol (equilibrated to room temperature). After extraction, the phases are separated in a microfuge (15,000 rpm for 3 minutes at room temperature) and the phenolic phase is removed. The aqueous phase and interface material are re-extracted with water-saturated phenol until little or no precipitate remains at the interface. The aqueous phase is then removed and the 2.9 kb DNA fragment is concentrated by ethanol precipitation.

One-half microgram of Bsu 36I/Sse 8387I linearized 6.2.1 viral DNA (see Example 9) is mixed with approximately 12 ng of the Bsu 36I/Sse 8387I fragment of NW44.1 in a 5 μl reaction mixture containing 25 mM Tris-HCl (pH 7.6), 5 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 5% (w/v) polyethylene glycol-8000, and 0.5 units T4 DNA ligase (Gibco-BRL, Gaithersburg, MD). After an overnight incubation at 16°C the entire ligation reaction is used to transfect Sf9 cells.

For transfection, 1.5 x 10^6 Sf9 cells are plated in one well of a 6-well cluster dish. After the cells have attached, the cell culture medium is replaced with 0.375 ml Grace's Insect cell culture medium (40). The contents of the ligation reaction are mixed with 0.375 ml of transfection buffer (25 mM HEPES (pH 7.1), 140 mM NaCl, 125 mM $CaCl_2$) and then added dropwise to the plated cells. A separate well is similarly treated with one half microgram of linearized (unligated) 6.2.1 viral DNA to provide a

negative control. The cells are incubated with the DNA for 4 hours at 27°C, washed once with 2 ml of Grace's Insect medium supplemented with 0.33% (w/v) lactalbumin hydrolysate, 0.33% (w/v) TC yeastolate, 0.1% (v/v) Pluronic™ F-68 (Gibco BRL, Gaithersburg, MD) and 10% (v/v) fetal bovine serum (complete TNM-FH medium), and then incubated for 2 hours at 27°C with 2 ml of complete TNM-FH. The cells are then harvested and one-tenth of the total culture (approximately 150,000 transfected cells) is mixed with 2 x 10^6 untreated Sf9 cells. The mixture is plated on a 6 cm tissue culture dish and the attached cells are carefully overlaid with 4 ml complete TNM-FH medium supplemented with 1.5% SeaPlaque™ agarose (FMC BioProducts, Rockland, ME), 100 units/ml penicillin G and 100 μg/ml streptomycin sulfate. Since the cells are harvested 6 hours after transfection (i.e., before the production of extracellular virus), viral plaques are produced only by those cells which have taken up infectious viral DNA. Hence, the number of plaques on each plate provides a direct measure of the efficiency of the transfection event.

Five days after plating, the dishes are scanned for the presence of occ⁺ plaques. For one experiment, fifteen plaques are observed on a plate containing linearized and unligated 6.2.1 viral DNA (the negative control). Sixty-nine plaques are observed on the plate containing 6.2.1 viral DNA ligated to the NW44.1 Bsu 36I/Sse 8387I fragment. Eighteen of these plaques are picked at random for further analysis.

The plaques are transferred to individual wells of a 48-well cluster dish containing 7.5 x 10^4 Sf9 cells and 0.5 ml complete TNM-FH media. After 5 days, the extracellular virus is harvested from the

wells and analyzed by PCR (see Example 2) for the presence of the NW44.1-derived Bluescript sequences in the viral genome. The primers used for PCR are PVLReverse (see Figure 4), which anneals to the viral DNA approximately 320 bp downstream of the site of insertion of the Bsu-Sse linker in 6.2.1 (Example 2), and Bluescript "sequencing primer", which corresponds to the sequence on Bluescript SK+ DNA approximately 100 bp upstream of the insertion of the Bsu-Sse linker:

Bluescript sequencing primer: 5'-CCATGATTACGCCAAGCGCG-3' (SEQ ID NO:31)

With this primer set, a recombinant virus containing the NW44.1-derived Bluescript sequences yields a PCR product of approximately 400 bp in length, whereas no specific PCR products are formed with non-recombinant 6.2.1 viral DNA. The conditions for PCR are as described in Example 2. One-fifth of the PCR reaction is analyzed on a 1.8% agarose gel. All of the test samples contain the predicted 400 bp amplification product, indicating that each of the eighteen randomly picked viruses contains the desired insert. This result not only demonstrates the feasibility of the direct ligation approach, but also shows that the efficiency of recombinant virus recovery is very high.

Example 11

Construction of Biologically Active Recombinant AaIT-Expressing Viruses by Direct Ligation

To demonstrate that the direct ligation approach can be used to produce biologically active recombinant viruses, Bsu 36I/Sse 8783I delineated

virus insertion modules are isolated from six AaIT-containing modular expression vectors: pMEV1/codon optimized cuticle signal and codon optimized AaIT gene (Cuticle-AaIT), pMEV2/Cuticle-AaIT, pMEV3/Cuticle-AaIT, pMEV1/native AaIT signal and gene (AaIT-cDNA), pMEV2/AaIT-cDNA and pMEV3/AaIT-cDNA. An aliquot of each module is ligated with an appropriate amount of Bsu 36I/Sse 8387I linearized and purified AcMNPV A4000 DNA and transfected into Sf9 cells, as described in Example 10. In a separate experiment, two of the virus insertion modules (from pMEV1/Cuticle-AaIT and pMEV1/AaIT-cDNA) are ligated with linearized AcMNPV 6.2.1 viral DNA instead of A4000 viral DNA and transfected into Sf9 cells.

Linearized A4000 and 6.2.1 viral DNAs treated with DNA ligase in the absence of a virus insertion module are used as negative controls for the transfection. Five days after transfection, the medium is removed from the transfected Sf9 cells. Ten-fold serial dilutions of each transfection supernatant are prepared and two 1 ml aliquots of the 10^{-3}, 10^{-4} and 10^{-5} dilutions are used to infect 1.5 x 10^{6} cells in each of six 60 mm culture dishes. One hour after addition of virus, the virus inocula are removed and the cells are overlaid with agarose-containing medium, as described in the previous example. The titer of each transfection supernatant is calculated by counting viral plaques, and 12 or more plaques are picked at random from the 10^{-4} or 10^{-5} dilutions. Each plaque is then screened by PCR for the presence of the desired insert using the general procedure described in Example 2. In each case, the primer set is specific for the desired recombinant virus. The virus titers of the transfection supernatants (5 days post-transfection) and the

frequency of recombinant virus recovery is summarized below (pfu are plaque forming units):

Summary of Recombinant Virus Formation By Direct Ligation		
Derivatives of AcMNPV A4000		
Virus Insertion Module	Titer (pfu/ml)	No. PCR Positive/No. Tested
None	0.71×10^6	-
pMEV1/Cuticle-AaIT	1.4×10^6	2/12
pMEV2/Cuticle-AaIT	1.6×10^6	5/16
pMEV3/Cuticle-AaIT	1.8×10^6	5/16
pMEV1/AaIT-cDNA	1.4×10^6	4/12
pMEV2/AaIT-cDNA	0.81×10^6	0/40
pMEV3/AaIT-cDNA	1.4×10^6	3/12
Derivatives of AcMNPV 6.2.1		
None	1.1×10^6	-
pMEV1/Cuticle-AaIT	1.7×10^6	1/12
pMEV1/AaIT-cDNA	1.3×10^6	3/12

One virus (A4000 containing the pMEV2/AaIT cDNA module) is not recovered on this initial screening. This virus appears to replicate poorly and is recovered from a primary screening of a transfection supernatant isolated 2 days, rather than 5 days, after DNA addition. The frequency with which recombinant viruses are recovered from the other transfection supernatants ranges from 8%-33%.

One to three recombinants identified on the

first round of screening for each virus are then subjected to one or two additional rounds of plaque purification, and P1 stocks are prepared as described in Example 2. Since the AcMNPV A4000 derivatives do not form orally infectious occlusion bodies, their biological activity is assessed by injecting 6,000-10,000 pfu of extracellular virus into the body cavity of fourth instar Heliothis virescens (tobacco budworm) larvae. The virus is titered by the plaque assay method, and then diluted to 1.2-2 x 10^7 pfu/ml in TNM-FH medium supplemented with 0.5% (v/v) red dye number 5. Each larva is anesthetized with carbon dioxide for 2-5 minutes and then injected with 0.5 μl of diluted virus, using a Hamilton syringe equipped with a 26 gauge needle. The needle is inserted longitudinally between the last two prolegs and then moved anteriorly two to three body segments prior to injection. Following injection, each larva is inspected for the release of dye-stained hemolymph and discarded if sample loss is evident or suspected. The larvae are then maintained at 27°C in covered 4 cm^2 diet cells (one larva per cell) and inspected visually 1-4 times a day for evidence of morbidity or mortality. An infected larva is scored as responding to the treatment if it is either dead or moribund. A moribund individual is one which is unable to right itself within 0.5-2 minutes after being turned on its back.

For the two 6.2.1 derivatives, which do produce orally infectious occlusion bodies, biological activity is assessed on third instar Heliothis virescens larvae following oral administration of 1000 occlusion bodies on a small leaf disk. A one μl droplet containing the virus in TET buffer (50 mM Tris-HCl, pH 7.5/10 mM EDTA/0.1% Triton™ X-100 (Rohm

and Haas, Philadelphia, PA)) is added to a 5 mm diameter cotton leaf disk in an individual well containing a premoistened filter paper disk (approximately 4 mm in diameter - 30 μl water). After the droplet dries, a single larva is placed in the well and the well is closed. The larvae are allowed to feed overnight. The next morning, larvae which consume the entire leaf disk are transferred to individual wells containing insect diet. The larvae are monitored for mortality until survivors pupate. This procedure is also performed with a wild-type E2 AcMNPV viral strain, which is used as a control.

In both bioassays, greater than 95% of all responding larvae infected with the recombinant virus containing the AaIT gene exhibit contractile paralysis prior to death. Moreover, as summarized below, all of the recombinant AaIT viruses have a shorter mean response time (RT_{50}) than wild type AcMNPV (strain E2). If more than one virus isolate is tested in each virus group, a separate RT_{50} value is reported for each isolate.

Bioassay Summary: Recombinant AaIT Viruses on *H. virescens*			
Derivatives of AcMNPV A4000			
AaIT gene	Expression Vector	Promoter	RT50 % of wild type
Cuticle-AaIT	pMEV1	DA26	50± 4 52± 5
Cuticle-AaIT	pMEV2	6.9K	51± 2 51± 3
Cuticle-AaIT	pMEV3	Polyhedrin	53 64± 11
AaIT-cDNA	pMEV1	DA26	69± 3
AaIT-cDNA	pMEV2	6.9K	51± 3
AaIT-cDNA	pMEV3	Polyhedrin	57± 2
Derivatives of AcMNPV 6.2.1			
Cuticle-AaIT	pMEV1	DA26	62
AaIT-cDNA	pMEV1	DA26	72

Samples of an isolate designated A4001 (containing the cuticle/AaIT gene under the control of the DA26 promoter inserted into the A4000 direct ligation virus vector) and of an isolate designated A1001 (containing the cuticle/AaIT gene under the control of the DA26 promoter inserted into the 6.2.1 direct ligation virus vector) have been deposited by applicants with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., on April 7, 1993, and have been assigned ATCC accession numbers VR-2405 (for A4001) and VR-2404 (for A1001).

Bibliography

1. Smith, G. E., and Summers, M. D., U.S. Patent Number 4,745,051.

2. Francki, R. I. B., et al., eds., "Classification and Nomenclature of Viruses" in Archives of Virology, Supp. 2, pages 117-123 (1991).

3. Luckow, V. A., and Summers, M. D., Bio/Technology, 6, 47-53 (1988).

4. Webb, N. R., and Summers, M. D., Technique, 2, 173-188 (1990).

5. Kitts, P. A., et al., Nucl. Acids Res., 18, 5667-5672 (1990).

6. Kitts, P. A., and Possee, R. D., BioTechniques, 14, 810-817 (1993).

7. Lerch, R. A., and Friesen, P. D., Nucl. Acids Res., 21, 1753-1769 (1993).

8. Patel, G., et al., Nucl. Acids Res., 20, 87-104 (1992).

9. Peakman, T. C., et al., Nucl. Acids Res., 20, 495-500 (1992).

10. Summers, M. D., and Smith, G. E., A Manual Of Methods For Baculovirus Vectors and Insect Cell Culture Procedures, Dept. of Entomology, Texas Agricultural Experiment Station and Texas A & M University, College Station, Texas 77843-2475, Texas Agricultural Experiment Station Bulletin No. 1555 (1987).

11. Granados, R. R., and Federici, B. A., The Biology of Baculoviruses, I, 99 (1986).

12. Published Australian patent application number 35291/89

13. Kuzio, J., et al., Virology, 173, 759-763 (1989).

14. Tomalski, M. D., and Miller, L. K.,

Nature, 352, 82-85 (1991).

15. Zlotkin, E., et al., Toxicon, 9, 1-8 (1971).

16. Jackson, J. R. H., and Parks, T. N., United States Patent Number 4,925,664.

17. Martens, J. W. M., et al., App. & Envir. Microbiology, 56, 2764-2770 (1990).

18. Federici, B. A., In Vitro, 28, 50A (1992).

19. Eldridge, R., et al., Insect Biochem., 21, 341-351 (1992).

20. Menn, J. J., and Borkovec, A. B., J. Agric. Food Chem., 37, 271-278 (1989).

21. Hammock, B. D., et al., Nature, 344, 458-461 (1990).

22. Guarino, L. A., et al., J. Virology, 60, 224-229 (1986).

23. Sambrook, J., et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

24. Emery, and Bishop, Protein Engineering, 1, 359-366 (1987).

25. Wang, et al., Gene, 100, 131-137 (1991).

26. Luckow, V. A., and Summers, M. D., Virology, 170, 31-39 (1989).

27. Malitschek, B., and Schartl. M., BioTechniques, 11, 177-178 (1991).

28. Wilson, M. E., et al., J. Virology, 61, 661-666 (1991).

29. Hill-Perkins, M. S., and Possee, R. D., J. Gen. Virology, 71, 971-976 (1990).

30. O'Reilly, D. R., et al., J. Gen. Virology, 71, 1029-1037 (1990).

31. Friesen, P.D., and Miller., L. K., J. Gen. Virology, 71, 2264-2272 (1987).

32. Morris, and Miller, J. Virology, 66, 7397-7405 (1992).

33. Cochran, M., and Faulkner, P., J. Virology, 45, 961-970 (1983).

34. Guarino L., and Dong, W., J. Virology, 65, 3676-3680 (1991).

35. Guarino, L. A., et al., J. Virology, 60, 215-223 (1986).

36. Pearson, M. R., et al., Science, 257, 1382-1384 (1992).

37. Stewart, L. M. D., et al., Nature, 352, 85-88 (1991).

38. Maeda, S., Biochem. Biophys. Res. Commun., 165, 1177-1183 (1989).

39. Bougis, P. E., et al., J. Biol. Chem., 264, 19259-19265 (1989).

40. Grace, T. D. C., Nature, 195, 788-789 (1962).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: American Cyanamid Company

(ii) TITLE OF INVENTION: Gene Insertion By Direct Ligation In Vitro

(iii) NUMBER OF SEQUENCES: 45

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: American Cyanamid Company
(B) STREET: One Cyanamid Plaza
(C) CITY: Wayne
(D) STATE: New Jersey
(E) COUNTRY: U.S.A.
(F) ZIP: 07470-8426

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT/US94/
(B) FILING DATE: 27-MAY-1994
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Gordon, Alan M
(B) REGISTRATION NUMBER: 30,637
(C) REFERENCE/DOCKET NUMBER: 31969-00/PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 201-831-3244
(B) TELEFAX: 201-831-3305

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCTCAGGGCA GCTTAAGGCA GCGGACCGGC AGCCTGCAGG 40

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCTGCAGGCT GCCGGTCCGC TGCCTTAAGC TGCCCTGAGG 40

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGATTTCCTT GAAGAGAGTG AG 22

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 785 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGGCGGTTT TAACAGCCGT CGATTTAACT AATGCCAGTA GGTATGCCAT ACATATGCAT 60

CGTCTCGAGG TCCCCTGCAG GCTGCCGGTC CGCTGCCTTA AGCTGCCCTG AGGACGGTAT 120

CGATAAGCTT GATAACTCGA ATCGCTATCC AAGCCAGCTC CATTGTCGGC ATCGTGCTCA 180

TTCTATTGAC GCTGGCAGAT TTGGTTTTGG CGCTATGGGA CCCGTTCGGT TACAACAACA 240

TGTTTCCGCG CGAGTTTCCC GACGACATGT CGCGCACGTT CCTGACTGCG TACTTTGAGA 300

GTTTCGACAA CACCACGTCC AGAGAAATCA TAGAGTTTAT GCCCGAGTTC TTTTCGGAAA 360

TGGTCGAAAC GGACGATGAC GCCACGTTTG AATCTCTATT TCATTTATTA GATTATGTGG 420

CATCTTTAGA AGTTAATTCC GACGGCCAAA TGTTAAACTT GGAGGAGGGT GATGAAATTG 480

AGGATTTTGA CGAATCTACT TTGGTGGGGC AAGCGTTAGC CACTAGCTCG CTATACACTC 540

GCATGGAGTT TATGCAGTAC ACGTTTAGGC AAAACACACT ATTGTCTATG AACAAAGAAA 600

ACAACAATTT TAATCAAATA ATCATGGGTT TATTTGCAAC AAACACAATT GTGGCGTTTA 660

CAGCATTTGT TATACACACA GAACTCATAT TTTTTATATT TTTCGTAATC TTCCTAATGA 720

TCACATTTTA TTACATAATC AAAGAATCGT ACGAATATTA TAAAACAATT GATTTGTTAT 780

TTTAA 785

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

NNNNNNNNNN NNNNNNNGCA GGT 23

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGCAGCCTCG AGCCTCAGGC CTATGCCGTG TCCAATTGCA AG 42

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGACAGGATC CATTACCTGC AGGAGTTTAA ATTGTGTAAT TTATGTAG 48

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 248 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGCAGCCTCG AGCCTCAGGC CTATGCCGTG TCCAATTGCA AGTTCAACAT TGAGGATTAC 60

AATAACATAT TTAAGGTGAT GGAAAATATT AGGAAACACA GCAACAAAAA TTCAAACGAC 120

CAAGACGAGT TAAACATATA TTTGGGAGTT CAGTCGTCGA ATGCAAAGCG TAAAAAATAT 180

TAATAAGGTA AAAATTACAG CTACATAAAT TACACAATTT AAACTCCTGC AGGTAATGGA 240

TCCTGTCC 248

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GATCTAGAAT TCCATGGATC CCGGGTACCA ACCAGACATT CCACACAGC 49

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGCAGCGAGC TCCTGCAGGC CTCAAACACA GGCAAATATT GA 42

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 242 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GATCTAGAAT TCCATGGATC CCGGGTACCA ACCAGACATT CCACACAGCC GACAGTAGCG 60

AATGAACGAA GCGATTTCGT CGCCTGCCCT CGTTTGGCTT TCGACTGTTA CAAAATCATG 120

TCTGCAAGAT TTTAAAYTAA GCCCGCTAAG CTCAAATAGT TTATTTTTAT TACTGTTTTG 180

TAAATAAATA ACTTTATCAT TCAATATTTG CCTGTGTTTG AGGCCTGCAG GAGCTCGCTG 240

CT 242

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

AGCAGCGAGC TCCTGCAGGC CTACGCGTAA TTCGATATAG AC 42

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGGATCTAGA CACGTCTCGT TCGATGTTTC GCCTTTGAAC GT 42

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGCAGCGAGC TCCTGCAGGC CTACGCGTAA TTCGATATAG ACATGACATC AGTCGTCAAT 60

TGTATTCAAA AACAACAATT GCTGCCAATG TACCGTATTC AAATTACTAC ATGTATAAAT 120

CTGTGTTTTC TATTGTAATG AATCACTTAA CACACTTTTA ATTACGTCAA TAAATGTTAT 180

TCACCATTAT TTACCTGGTT TTTTTGAGAG GGGCTTTGTG CGACTGCGCA CTTCCAGCCT 240

TTATAAACGC TCACCAACCA AAGCAGGTCA TTATTGTGCC AGGACGTTCA AAGGCGAAAC 300

ATCGAACGAG ACGTGTCTAG ATCCG 325

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AGCAGCGAGC TCCTGCAGGC CTCTTGATGT CTCCGATTTC 40

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CGGATCTAGA CACGTCTCGT TTGCTATGGT AAAGCTCAAA 40

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 440 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AGCAGCGAGC TCCTGCAGGC CTCTTGACGT CTCCGATTTC TTTTTGGCGG CAATAAGCAC 60

TCCAATGCAA ATACAAAACT TTGTCGCAAC TACTGATGTT TTCGATTTCA TTCTGAAATT 120

GTTCTAAAGT TTGTAACGCG TTCTTGTTAA AGTAATAGTC CGAGTTTGTC GACAAGGAAT 180

CGTCGGTGGC GTACACGTAG TAGTTAATCA TCTTGTTGAT TGATATTTAA TTTTGGCGAC 240

GGATTTTTAT ATACACGAGC GGAGCGGTCA CGTTCTGTAA CATGAGTGAT CGTGTGTGTG 300

TTATCTCTGG CAGCGCGATA GTGGTCGCGA AAATTACACG CGCGTCGTAA CGTGAACGTT 360

TATATTATAA ATATTCAACG TTGCTTGTAT TAAGTGAGCA TTTGAGCTTT ACCATAGCAA 420

ACGAGACGTG TCTAGATCCG 440

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AGCAGCGAGC TCCTGCAGGC CTATGCCGTG TCCAATTGCA AG 42

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CGGATCTAGA CACGTCTCGG TTTAAATTGT GTAATTTATG TA 42

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 243 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGCAGCGAGC TCCTGCAGGC CTATGCCGTG TCCAATTGCA AGTTCAACAT TGAGGATTAC 60

AATAACATAT TTAAGGTGAT GGAAAATATT AGGAAACACA GCAACAAAAA TTCAAACGAC 120

CAAGACGAGT TAAACATATA TTTGGGAGTT CAGTCGTCGA ATGCAAAGCG TAAAAAATAT 180

TAATAAGGTA AAAATTACAG CTACATAAAT TACACAATTT AAACCGAGAC GTGTCTAGAT 240

CCG 243

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AGCAGCGAGC TCCTGCAGGC CTGACGCACA AACTAATATC AC 42

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CGGATCTAGA CACGTCTCGA TTTATAGGTT TTTTTATTAC AA 42

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 188 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AGCAGCGAGC TCCTGCAGGC CTGACGCACA AACTAATATC ACAAACTGGA AATGTCTATC 60

AATATATAGT TGCTGATATC ATGGAGATAA TTAAAATGAT AACCATCTCG CAAATAAATA 120

AGTATTTTAC TGTTTTCGTA ACAGTTTTGT AATAAAAAAA CCTATAAATC GAGACGTGTC 180

TAGATCCG 188

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 847 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AGCAGCGAGC TCCTGCAGGC CTGTCTTGGC TTTCCTCATA TGTACATATG TATGTAAATA 60

TGTAAAATAA GTCGCAACTA AATTCTAATA CATTTTTCAG AATCTTAAAT TAATTTTATC 120

GTATATTAAA ACAGAAGAAA GTCCGTTAAT AGTTGATTTC ATTAACTAAA AGTACAAAAT 180

AATCTTTAAT ACATATGCCG ATCAGACATT TATTGGTTTA GAAGCGCAGT ATTTTTTTTG 240

CGAATACGCA TAACAAAGCG CTTCGATTAT CTTTAACATA AGTTATTTAA GCAGCCGTAT 300

TTATAAAGAA ATTTCCAAAA TAAAGCGAAT ATTCTAGAAT CCCAAAACAA ACTGGTTATT 360

GTGGTAGGTC ATTTGTTTGG CAGAAAGAAA ACTCGAGAAA TTTCTCTGGC CGTTATTCGT 420

TATTCTCTCT TTTCTTTTTG GGTCTCTCCC TCTCTGCACT AATGCTCTCT CACTCTGTCA 480

CACAGTAAAC GGCATACTGC TCTCGTTGGT TCGAGAGAGC GCGCCTCGAA TGTTCGCGAA 540

AAGAGCGCCG GAGTATAAAT AGAGGCGCTT CGTCTACGGA GCGACAATTC AATTCAAACA 600

AGCAAAGTGA ACACGTCGCT AAGCGAAAGC TAAGCAAATA AACAAGCGCA GCTGAACAAG 660

CTAAACAATC TGCAGTAAAG TGCAAGTTAA AGTGAATCAA TTAAAAGTAA CCAGCAACCA 720

60

AGTAAATCAA CTGCAACTAC TGAAATCTGC CAAGAAGTAA TTATTGAATA CAAGAAGAGA 780

ACTCTGAATA CTTTCAACAA GTTACCGAGA AAGAAGAACT CACACACACG AGACGTGTCT 840

AGATCCG 847

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATTCCATGGA TCCCGGGTAC CTGTAACTAG TGCACTCAAC 40

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AGCAGCCTCG AGCCTCAGGC CTCCACATTG TCGACTTGCT 40

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATTCCATGGA TCCCGGGTAC CTGTAACTAG TGCACTCAAC AAAAATGTAA TATTAAACAC 60

AATTAAATAA ATGTTGAAAA TTTATTGCCT AATATTATTT TTGTCAGTTS STTGTCATTT 120

ATTAATTTGG ATGATGTCCA TTTGTTTTTA AAATTGAACT GGCTTTACGA GTAGAATTCT 180

ACGCGTAAAA CACAATCAAG TAYGAGTCAT AAGCTGATGT CATGTTTTGC ACACGGCTCA 240

TAACCGAACT GGCTTTACGA GTAGAATTCT ACTTGTAACG CACGATCGAG TGGATGATGG 300

TCATTTGTTT TTCAAATCGA GATGATGTCA TGTTTTGCAC ACGGGCTCAT AAACTCGCTT 360

TACGAGTAGA ATTCTACGTG TAACGCACGA TCGATTGATG AGTCATTTGT TTTGCAATAT 420

GATATCATAC AATATGACTC ATTTGTTTTT CAAAACCGAA CTTGATTTAC GGGTAGAATT 480

CTACTYGTAA AGCACAATCA AAAAGATGAT GTCATTTGTT TTTCAAAACT GAACTCTCGG 540

CTTTACGAGT AGAATTCTAC GTGTAAAACA CAATCAAGAA ATGATGTCAT TTGTTATAAA 600

AATAAAAGCT GATGTCATGT TTTGCACATG GCTCATAACT AAACTCGCTT TACAAATAGA 660

ATTCTACGCG TAAAACATGA TTGATAATTA AATAATTCAT TTGCAAAGCT ATACGTTAAA 720

TCAAACGGAC GTTATGGAAT TGTATAATAT TAAATATGCA ATTGATCCAA CAAATAAAAT 780

TRTAATAGAG CAAGTCGACA ATGTGGAGGC CTGAGGCTCG AGGCTGCT 828

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 54 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ATGAAATTTC TCCTATTGTT TCTCGTAGTC CTTCCAATAA TGGGGGTGCT TGGC 54

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

ATGTTCAAGT TCGTGATGAT CTGCGCCGTC CTCGGCCTGG CTGTGGCCAA GAAGAACGGC 60

TACGCAGTCG ACTCATCCGG AAAAGCCCCC GAGTGCCTGC TCTCGAACTA TTGCAACAAT 120

GAATGCACCA AGGTGCACTA CGCTGACAAG GGCTACTGTT GCCTTCTGTC CTGCTATTGC 180

TTCGGTCTCA ACGACGACAA GAAAGTTCTG GAAATCTCTG ATACTCGCAA GAGCTACTGT 240

GACACCACCA TCATTAACTA AGGATCCTTT CCTGGGACCC GGCAAGAACC AAAAACTCAC 300

TCTCTTCAAG GAAATCC 317

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 797 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GAGCTCCTGC AGGCCTACGC GTAATTCGAT ATAGACATGA CATCAGTCGT CAATTGTATT 60

CATTAAAAAC AACAGCTGCC AATGTACCGT ATTCAAATTA CTACATGTAT AAATCTGTGT 120

TTTCTATTGT AATGAATCAC TTAACACACT TTTAATTACG TCAATAAATG TTATTCACCA 180

TTATTTACCT GGTTTTTTTG AGAGGGGCTT TGTGCGACTG CGCACTTCCA GCCTTTATAA 240

ACGCTCACCA ACCAAAGCAG GTCATTATTG TGCCAGGACG TTCAAAGGCG AAACATCGAA 300

ATGTTCAAGT TCGTGATGAT CTGCGCCGTC CTCGGCCTGG CTGTGGCCAA GAAGAACGGC 360

TACGCAGTCG ACTCATCCGG AAAAGCCCCC GAGTGCCTGC TCTCGAACTA TTGCAACAAT 420

GAATGCACCA AGGTGCACTA CGCTGACAAG GGCTACTGTT GCCTTCTGTC CTGCTATTGC 480

TTCGGTCTCA ACGACGACAA GAAAGTTCTG GAAATCTCTG ATACTCGCAA GAGCTACTGT 540

GACACCACCA TCATTAACTA AGGATCCCGG GTACCAACCA GACATTCCAC ACAGCCGACA 600

GTAGCGAATG AACGAAGCGA TTTCGTCGCC TGCCCTCGTT TGGCTTTCGA CTGTTACAAA 660

ATCATGTCTG CAAGATTTTA AACTAAGCCC GCTAAGCTCA AATAGTTTAT TTTTATTACT 720

GTTTTGTAAA TAAATAACTT TATCATTCAA TATTTGCCTG TGTTTGAGGC CTGAGGCTCG 780

AGGGGGGGCC CGGTACC 797

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CCATGATTAC GCCAAGCGCG 20

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ATGAAACTCC TGGTCGTGTT CGCCATGTGC GTGCCCGCTG CCAGCGCT 48

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATGAAACTTC TCGTTGTGTT CGCAATGTGC GTGCCTGCCG CCAGCGCC 48

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ATGTACAAAC TGACCGTCTT CCTGATGTTC ATCGCCTTCG TGATTATCGC TGAGGCC 57

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 57 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

ATGTACAAGC TCACAGTCTT CCTGATGTTC ATCGCTTTCG TCATCATCGC TGAGGCC 57

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ATGGCCGCTA AATTCGTCGT GGTTCTGGCC GCTTGCGTCG CCCTGAGCCA CTCGGCTATG 60

GTGCGCCGC 69

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

ATGGCAGCCA AGTTCGTCGT GGTTCTCGCC GCGTGCGTGG CCCTCTCGCA CAGCGCGATG 60

GTGCGCCGC 69

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ATGTTCACCT TCGCTATTCT GCTCCTGTGC GTGCAAGGCT GCCTGATCCA GAATGTTTAC 60

GGA 63

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

ATGTTTACCT TCGCTATTCT CCTTCTCTGC GTTCAGGGTT GCCTGATCCA AAATGTGTAC 60

GGT 63

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATGTTCAAGT TTGTCATGAT CTGCGCAGTT TTGGGCCTGG CGGTGGCC 48

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

ATGAACTACG TCGGGCTGGG CCTCATCATT GTGCTGTCGT GCTTGTGGCT GGGGAGCAAT 60

GCT 63

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 63 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

ATGAACTACG TGGGACTGGG ACTTATCATT GTGCTGAGCT GCCTTTGGCT CGGTTCGAAC 60

GCG 63

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

ATGCGCGTCC TGGTGCTGTT GGCCTGCCTG GCAGCCGCTA GCGCT 45

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

ATGAGGGTTC TAGTACTACT GGCCTGCTTG GCCGCGGCGT CAGCC 45

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 213 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AAGAAGAATG GATATGCCGT CGATAGTAGT GGTAAAGCTC CTGAATGTCT TTTGAGCAAT 60

TACTGTAACA ACGAATGCAC AAAAGTACAT TATGCTGACA AAGGATATTG CTGCTTACTT 120

TCATGTTATT GCTTCGGTCT AAATGACGAT AAAAAAGTTT TGGAGATTTC GGACACAAGG 180

AAAAGTTATT GTGACACCAC AATAATTAAT TAA 213

What is claimed is:

1. A recombinant double stranded DNA insect virus comprising:

(a) a double stranded DNA insect virus into which is inserted at least one recognition site for a restriction endonuclease which does not cut the viral genome to create a direct ligation virus vector; and

(b) a DNA fragment containing termini selected such that when the direct ligation virus vector of (a) is cleaved by the appropriate restriction endonuclease, the DNA fragment is ligated *in vitro* into the direct ligation virus vector of (a).

2. The recombinant double stranded DNA insect virus of Claim 1 wherein the virus is selected from the group consisting of double stranded enveloped DNA insect viruses from the Subfamilies *Entomopoxvirinae*, *Eubaculovirinae*, *Nudibaculovirinae*, *Ichnovirus* and *Bracovirus*, and double stranded nonenveloped DNA insect viruses from the family *Iridoviridae*.

3. The recombinant double stranded DNA insect virus of Claim 2 wherein the virus is the *Autographa californica* nuclear polyhedrosis virus (AcMNPV).

4. The recombinant double stranded DNA insect virus of Claim 1 wherein two recognition sites for a restriction endonuclease which does not cut the viral genome are inserted into the viral genome to create a direct ligation virus vector.

5. The recombinant double stranded DNA

insect virus of Claim 1, wherein the DNA fragment is inserted at any region of the viral genome which is nonessential for viral replication in cultured cells.

6. A modular expression vector which comprises a plasmid vector containing a virus insertion module which comprises in the following order:

(a) a recognition site for a restriction endonuclease;

(b) a promoter module containing a promoter and a 5′ untranslated region (UTR), where the 5′ UTR extends from the transcription start site to the last base pair which precedes the translation initiation codon for protein synthesis;

(c) a polylinker module to facilitate insertion of a heterologous gene;

(d) a 3′ UTR module containing at least a site for 3′ terminal mRNA processing and polyadenylation; and

(e) a recognition site for a restriction endonuclease,

such that the recognition sites of (a) and (e) permit the ligation in vitro of the virus insertion module into a direct ligation virus vector.

7. The modular expression vector of Claim 6 wherein: (a) the restriction endonuclease recognition site is selected from the group consisting of Bsp MI and Esp 3I recognition sites; (b) the promoter and 5′ UTR of the promoter module are selected from the group consisting of the heterologous promoter and 5′ UTR from the AcMNPV 6.9K gene, the AcMNPV DA26 gene, the AcMNPV polyhedrin gene, the AcMNPV 35K gene and the Drosophila melanogaster hsp70

gene; and (c) the 3' UTR module is selected from the group consisting of the (a) 3' UTR of the AcMNPV 6.9K gene and, (b) a region comprising the 3' terminus of the 35K gene together with the AcMNPV homologous region 5 (hr5).

8. The modular expression vector of Claim 6 wherein the polylinker module is altered by the insertion of a nucleic acid sequence encoding a heterologous protein.

9. The modular expression vector of Claim 8 wherein the toxin is AaIT.

10. The modular expression vector of Claim 8 wherein the nucleic acid sequence encoding a heterologous signal sequence is selected from the group consisting of the cuticle signal sequence from Drosophila melanogaster, the chorion signal sequence from Bombyx mori, the apolipophorin signal sequence from Manduca sexta, the sex specific signal sequence from Bombyx mori, the adipokinetic hormone signal sequence from Manduca sexta, the pBMHPC-12 signal sequence from Bombyx mori and the esterase-6 signal sequence from Drosophila melanogaster.

11. A recombinant double stranded DNA insect virus comprising:

(a) a double stranded DNA insect virus into which is inserted two recognition sites for a restriction endonuclease which does not cut the viral genome to create a direct ligation virus vector; and

(b) a DNA fragment containing termini selected such that when the direct ligation virus vector of (a) is cleaved by the appropriate restriction endonuclease, the DNA fragment is ligated in vitro into the direct

ligation virus vector of (a), wherein the DNA fragment comprises a virus insertion module which comprises in the following order:

(1) a recognition site for a restriction endonuclease;

(2) a promoter module containing a promoter and a 5' untranslated region (UTR), where the 5' UTR extends from the transcription start site to the last base pair which precedes the translation initiation codon for protein synthesis;

(3) a polylinker module altered by the insertion of a nucleic acid sequence encoding a heterologous protein;

(4) a 3' UTR module containing at least a site for 3' terminal mRNA processing and polyadenylation; and

(5) a recognition site for a restriction endonuclease.

12. A direct ligation virus vector comprising a double stranded DNA insect virus into which is inserted at least one recognition site for a restriction endonuclease which does not cut the viral genome, such that when the direct ligation virus vector is cleaved by the appropriate restriction endonuclease, a DNA fragment is ligated in vitro into the direct ligation virus vector.

13. The direct ligation virus vector of Claim 12 wherein the virus is selected from the group consisting of double stranded enveloped DNA insect viruses from the Subfamilies Entomopoxvirinae, Eubaculovirinae, Nudibaculovirinae, Ichnovirus and Bracovirus, and double stranded nonenveloped DNA

insect viruses from the family Iridoviridae.

14. The direct ligation virus vector of Claim 13 wherein the virus is the Autographa californica nuclear polyhedrosis virus (AcMNPV).

15. The direct ligation virus vector of Claim 12 wherein two recognition sites for a restriction endonuclease which does not cut the viral genome are inserted into the viral genome.

16. The direct ligation virus vector of Claim 12, wherein the DNA fragment is inserted at any region of the viral genome which is nonessential for viral replication in cultured cells.

Figure 1

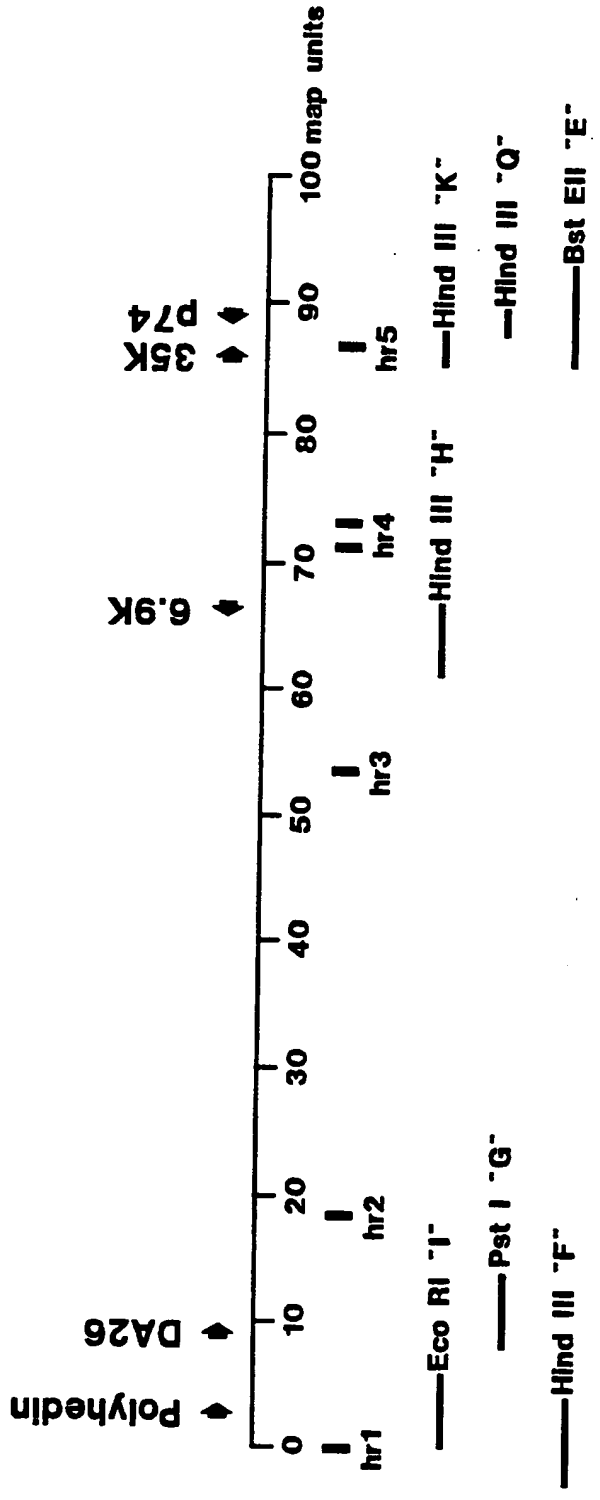

100 map units
90
80
70
60
50
40
30
20
10
0
hr1
hr2
hr3
hr4
hr5
Polyhedlin
DA26
6.9K
35K
P74
Eco RI "I"
Pst I "G"
Hind III "F"
Hind III "H"
Hind III "K"
Hind III "Q"
Bst EII "E"

Figure 2

Trinucleotide Frequencies in AcMNPV DNA (n = 30,866)				
	A	C	G	T
AA	0.0434	0.0228	0.0132	0.0300
AC	0.0219	0.0086	0.0183	0.0121
AG	0.0103	0.0119	0.0050	0.0120
AT	0.0196	0.0150	0.0158	0.0300
CA	0.0274	0.0131	0.0087	0.0166
CC	0.0115	0.0056	0.0099	0.0063
CG	0.0154	0.0152	0.0111	0.0201
CT	0.0092	0.0098	0.0103	0.0142
GA	0.0158	0.0113	0.0080	0.0140
GC	0.0157	0.0103	0.0161	0.0120
GG	0.0078	0.0116	0.0057	0.0098
GT	0.0139	0.0142	0.0146	0.0237
TA	0.0228	0.0136	0.0093	0.0197
TC	0.0167	0.0088	0.0176	0.0130
TG	0.0156	0.0154	0.0131	0.0244
TT	0.0228	0.0171	0.0279	0.0399

Figure 3

Tetranucleotide Frequencies in AcMNPV DNA (n=30,859)				
	A	C	G	T
AAA	0.01629	0.00946	0.00554	0.01215
AAC	0.00842	0.00239	0.00670	0.00534
AAG	0.00356	0.00388	0.00168	0.00411
AAT	0.00839	0.00563	0.00541	0.01059
ACA	0.00949	0.00414	0.00256	0.00570
ACC	0.00304	0.00162	0.00243	0.00152
ACG	0.00473	0.00505	0.00314	0.00541
ACT	0.00285	0.00294	0.00272	0.00362
AGA	0.00382	0.00223	0.00168	0.00265
AGC	0.00337	0.00210	0.00366	0.00278
AGG	0.00110	0.00168	0.00084	0.00139
AGT	0.00301	0.00226	0.00210	0.00469
ATA	0.00748	0.00346	0.00281	0.00586
ATC	0.00492	0.00200	0.00453	0.00359
ATG	0.00453	0.00317	0.00262	0.00547
ATT	0.00686	0.00440	0.00657	0.01218
CAA	0.01101	0.00573	0.00405	0.00667
CAC	0.00359	0.00230	0.00482	0.00239
CAG	0.00210	0.00324	0.00116	0.00226
CAT	0.00382	0.00353	0.00398	0.00531
CCA	0.00463	0.00256	0.00139	0.00298
CCC	0.00210	0.00077	0.00165	0.00106
CCG	0.00285	0.00265	0.00165	0.00281
CCT	0.00132	0.00174	0.00126	0.00197
CGA	0.00382	0.00408	0.00278	0.00473
CGC	0.00366	0.00353	0.00398	0.00411
CGG	0.00207	0.00398	0.00220	0.00291
CGT	0.00301	0.00512	0.00518	0.00686
CTA	0.00265	0.00236	0.00106	0.00314
CTC	0.00265	0.00191	0.00320	0.00207
CTG	0.00217	0.00262	0.00210	0.00340
CTT	0.00132	0.00288	0.00453	0.00544
GAA	0.00593	0.00372	0.00210	0.00411
GAC	0.00385	0.00197	0.00343	0.00210
GAG	0.00191	0.00230	0.00129	0.00249
GAT	0.00262	0.00291	0.00301	0.00554
GCA	0.00580	0.00366	0.00288	0.00343
GCC	0.00301	0.00184	0.00369	0.00181
GCG	0.00330	0.00421	0.00343	0.00518
GCT	0.00200	0.00275	0.00337	0.00395
GGA	0.00288	0.00194	0.00119	0.00178
GGC	0.00333	0.00197	0.00440	0.00197
GGG	0.00123	0.00191	0.00097	0.00168
GGT	0.00178	0.00233	0.00304	0.00268
GTA	0.00424	0.00395	0.00233	0.00340
GTC	0.00385	0.00243	0.00537	0.00259
GTG	0.00220	0.00359	0.00337	0.00550
GTT	0.00450	0.00392	0.00657	0.00868
TAA	0.01020	0.00395	0.00155	0.00709
TAC	0.00602	0.00194	0.00337	0.00233
TAG	0.00281	0.00249	0.00087	0.00320
TAT	0.00479	0.00294	0.00340	0.00861
TCA	0.00755	0.00275	0.00194	0.00453
TCC	0.00340	0.00136	0.00220	0.00191
TCG	0.00453	0.00337	0.00294	0.00677
TCT	0.00304	0.00239	0.00294	0.00466
TGA	0.00534	0.00311	0.00230	0.00492
TGC	0.00541	0.00275	0.00408	0.00320
TGG	0.00340	0.00411	0.00178	0.00385
TGT	0.00612	0.00453	0.00434	0.00946
TTA	0.00839	0.00388	0.00317	0.00735
TTC	0.00534	0.00252	0.00450	0.00479
TTG	0.00677	0.00605	0.00502	0.01011
TTT	0.01014	0.00596	0.01024	0.01364

Figure 4

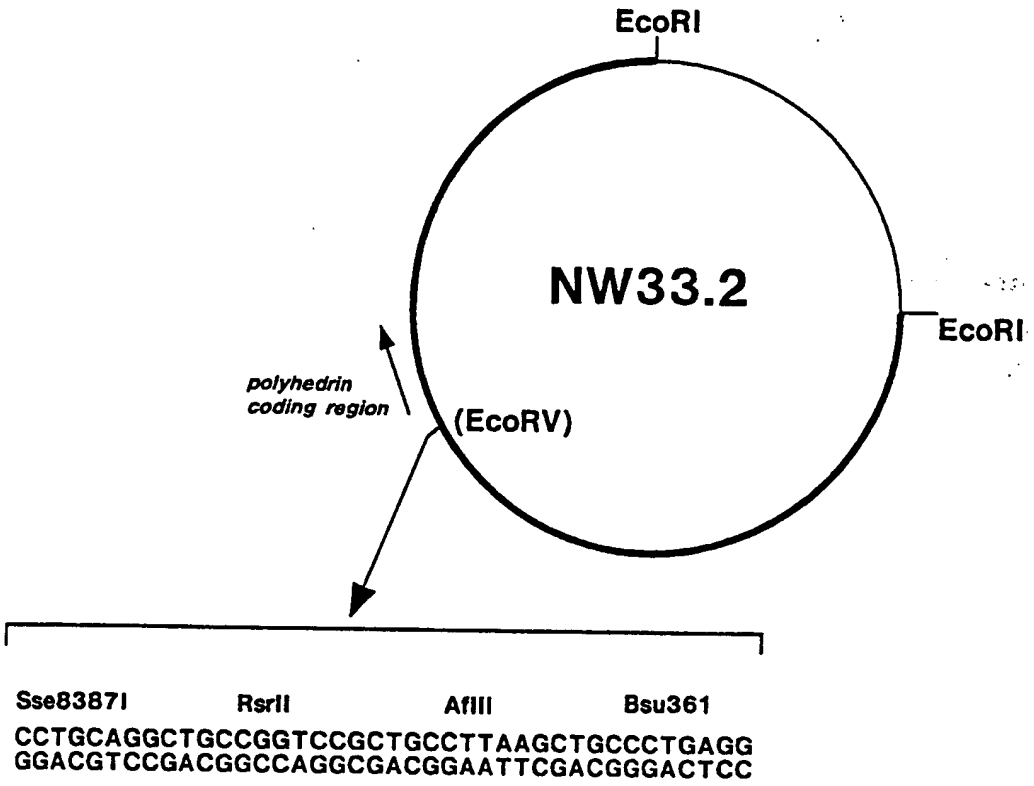
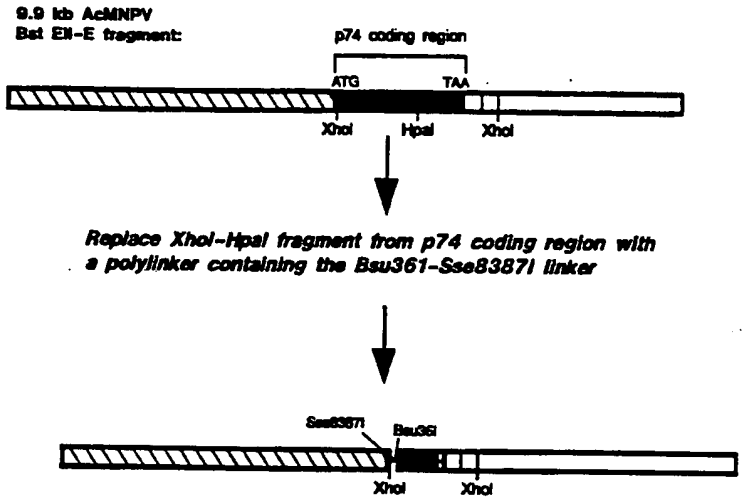

EcoRI
NW33.2
EcoRI
polyhedrin
coding region
(EcoRV)
Sse8387I
RsrII
AflII
Bsu361
CCTGCAGGCTGCCGGTCCGCTGCCTTAAGCTGCCCTGAGG
GGACGTCCGACGGCCAGGCGACGGAATTCGACGGGACTCC

Figure 5




```
+1                                                                    +69
I                                                                     I
ATGGCGGTTTTAACAGCCGTCGATTTAACTAATGCCAGTAGGTATGCCATACATATGCATCGTCTCGAG
                                                                   XhoI
      Sse8387I        RsrII          AflII          Bsu36I      ClaI   HindIII
GTCCCCTGCAGGCTGCCGGTCCGCTGCCTTAAGCTGCCCTGAGGACGGTATCGATAAGC
TTGATAACTCGAATCGCTATCCAAGCCAGCTCCATTGTCGGCATCGTGCTCATTCTATTGACGCTGGC
    +1287
AGATTTGGTTTTGGCGCTATGGGACCCGTTCGGTTACAACAACATGTTTCCGCGCGAGTTTCCCGACGA
CATGTCGCGCACGTTCCTGACTGCGTACTTTGAGAGTTTCGACAACACCACGTCCAGAGAAATCATAGA
GTTTATGCCCGAGTTCTTTTCGGAAATGGTCGAAACGGACGATGACGCCACGTTTGAATCTCTATTTCA
TTTATTAGATTATGTGGCATCTTTAGAAGTTAATTCCGACGGCCAAATGTTAAACTTGGAGGAGGGTGA
TGAAATTGAGGATTTTGACGAATCTACTTTGGTGGGGCAAGCGTTAGCCACTAGCTCGCTATACACTCG
CATGGAGTTTATGCAGTACACGTTTAGGCAAAACACACTATTGTCTATGAACAAAGAAAACAACAATTT
TAATCAAATAATCATGGGTTTATTTGCAACAAACACAATTGTGGCGTTTACAGCATTTGTTATACACAC
AGAACTCATATTTTTTATATTTTTCGTAATCTTCCTAATGATCACATTTTATTACATAATCAAAGAATC
                                                +1937
GTACGAATATTATAAAACAATTGATTTGTTATTTTAA
```

Figure 6

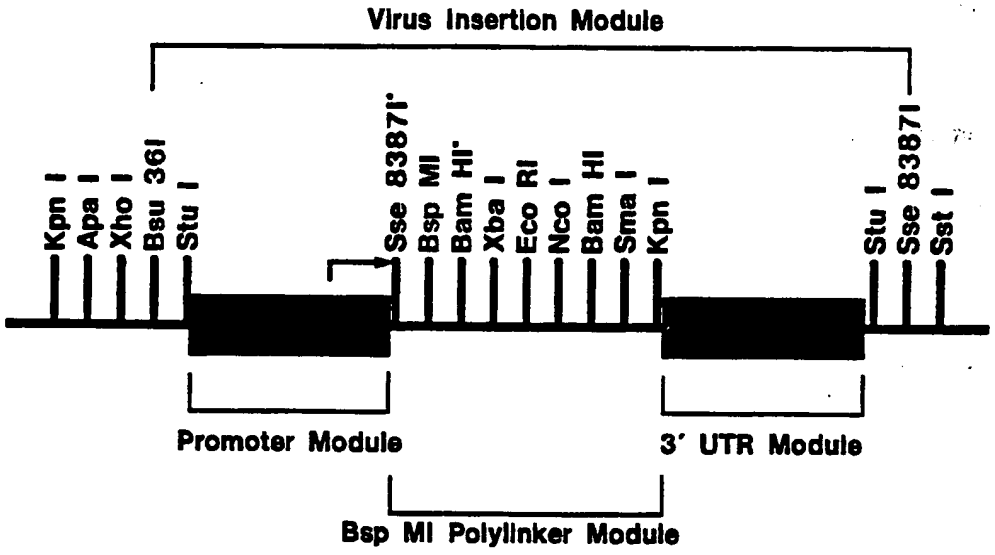

Virus Insertion Module
Kpn I
Apa I
Xho I
Bsu 36I
Stu I
Sse 8387I'
Bsp MI
Bam HI'
Xba I
Eco RI
Nco I
Bam HI
Sma I
Kpn I
Stu I
Sse 8387I
Sst I
Promoter Module
3' UTR Module
Bsp MI Polylinker Module

Figure 7

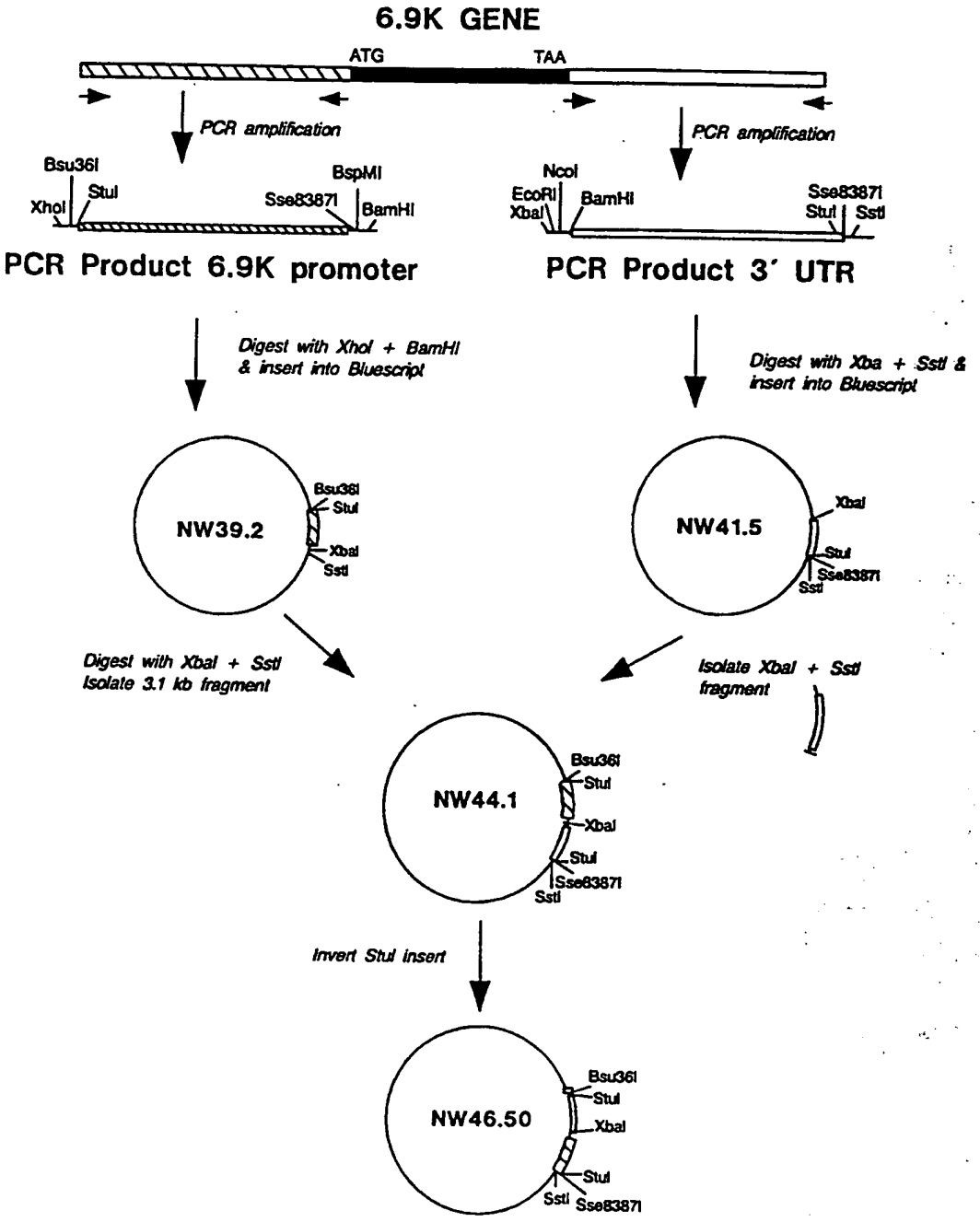

6.9K GENE
ATG
TAA
PCR amplification
PCR amplification
Bsu36I
XhoI
StuI
BspMI
Sse8387I
BamHI
NcoI
EcoRI
XbaI
BamHI
Sse8387I
StuI
SstI
PCR Product 6.9K promoter
PCR Product 3' UTR
Digest with XhoI + BamHI & insert into Bluescript
Digest with Xba + SstI & insert into Bluescript
NW39.2
Bsu36I
StuI
XbaI
SstI
NW41.5
XbaI
StuI
Sse8387I
SstI
Digest with XbaI + SstI Isolate 3.1 kb fragment
Isolate XbaI + SstI fragment
NW44.1
Bsu36I
StuI
XbaI
StuI
Sse8387I
SstI
Invert StuI insert
NW46.50
Bsu36I
StuI
XbaI
StuI
SstI
Sse8387I

Figure 8

PCR AMPLIFICATION OF THE AcMNPV 6.9K GENE PROMOTER FOR Bsp MI-BASED MODULAR EXPRESSION VECTORS

PRIMER SET:

Strand	Primer	Sequence
(+) strand	NW oligo 1	5' – AGCAGCCTCGAGCCTCAGGCCTATGCCGTGTCCAATTGCAAG – 3'
(–) strand	PD oligo 23	5' – GGACAGGATCCATTACCTGCAGGAGTTTAAATTGTGTAATTTATGTAG – 3'

TEMPLATE: AcMNPV Hind III fragment "H"

AMPLIFIED FRAGMENT:

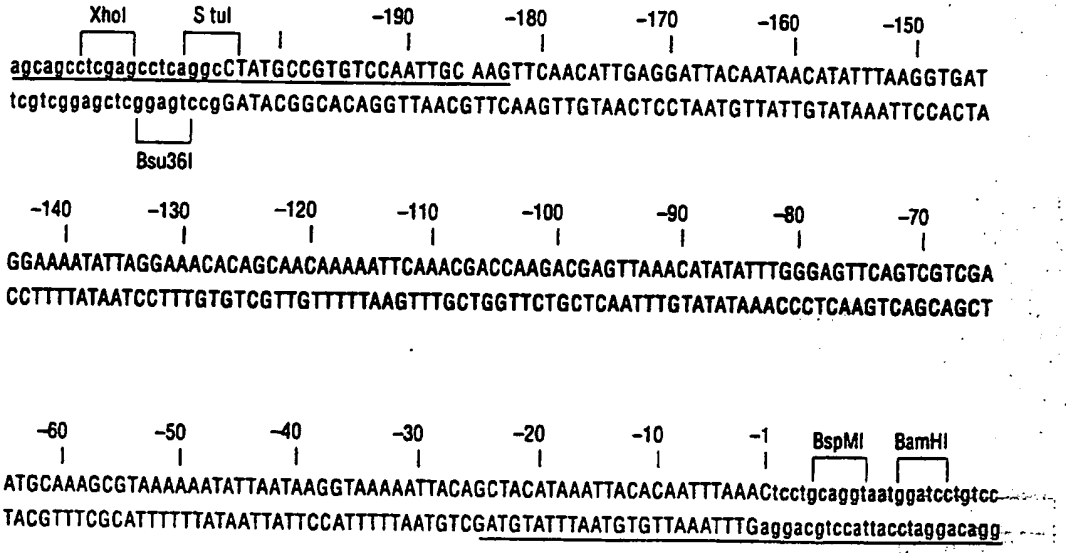


Figure 9

PCR AMPLIFICATION OF THE AcMNPV 6.9K GENE 3' - UTR FOR Bsp MI-BASED MODULAR EXPRESSION VECTORS

PRIMER SET:

Strand	Primer	Sequence
(+) strand	NW oligo 2	5' – GATCTAGAATTCCATGGATCCCGGGTACCAACCAGACATTCCACACAGC – 3'
(–) strand	NW oligo 3	5' – AGCAGCGAGCTCCTGCAGGCCTCAAACACAGGCAAATATTGA – 3'

TEMPLATE: AcMNPV Hind III fragment "H"

AMPLIFIED FRAGMENT:

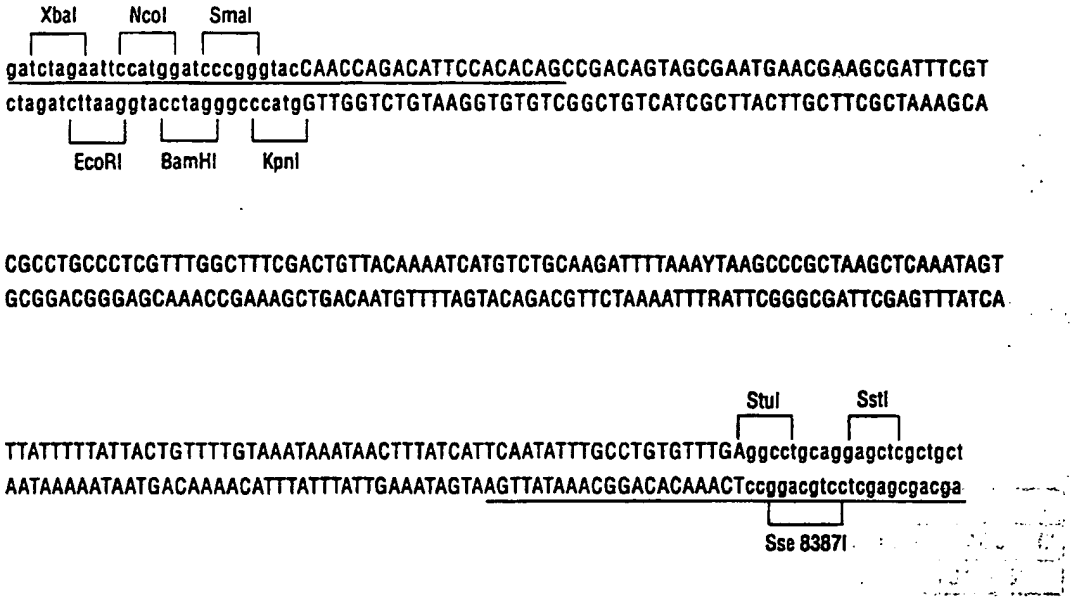


Figure 10

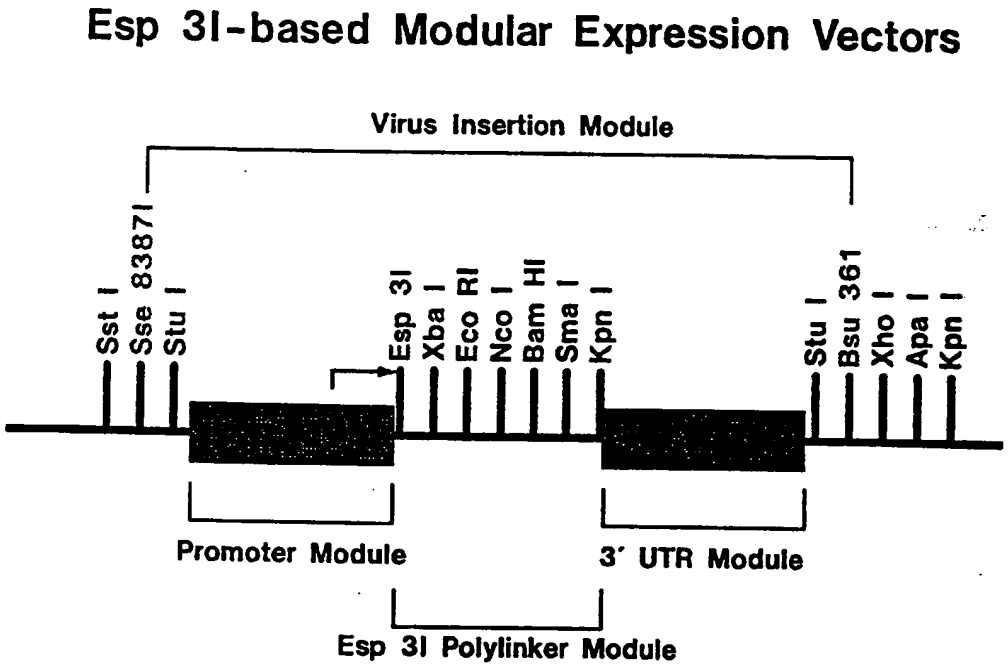

Esp 3I-based Modular Expression Vectors
Virus Insertion Module
Sst I
Sse 8387I
Stu I
Esp 3I
Xba I
Eco RI
Nco I
Bam HI
Sma I
Kpn I
Stu I
Bsu 361
Xho I
Apa I
Kpn I
Promoter Module
3' UTR Module
Esp 3I Polylinker Module

Figure 11

PCR AMPLIFICATION OF THE AcMNPV DA26 GENE PROMOTER FOR Esp 3I-BASED MODULAR EXPRESSION VECTORS

PRIMER SET:

Strand	Primer	Sequence
(+) strand	DA26FZ	5' – AGCAGCGAGCTCCTGCAGGCCTACGCGTAATTCGATATAGAC – 3'
(–) strand	DA26RZ	5' – CGGATCTAGACACGTCTCGTTCGATGTTTCGCCTTTGAACGT – 3'

TEMPLATE: AcMNPV Pst I fragment "G"

AMPLIFIED FRAGMENT:

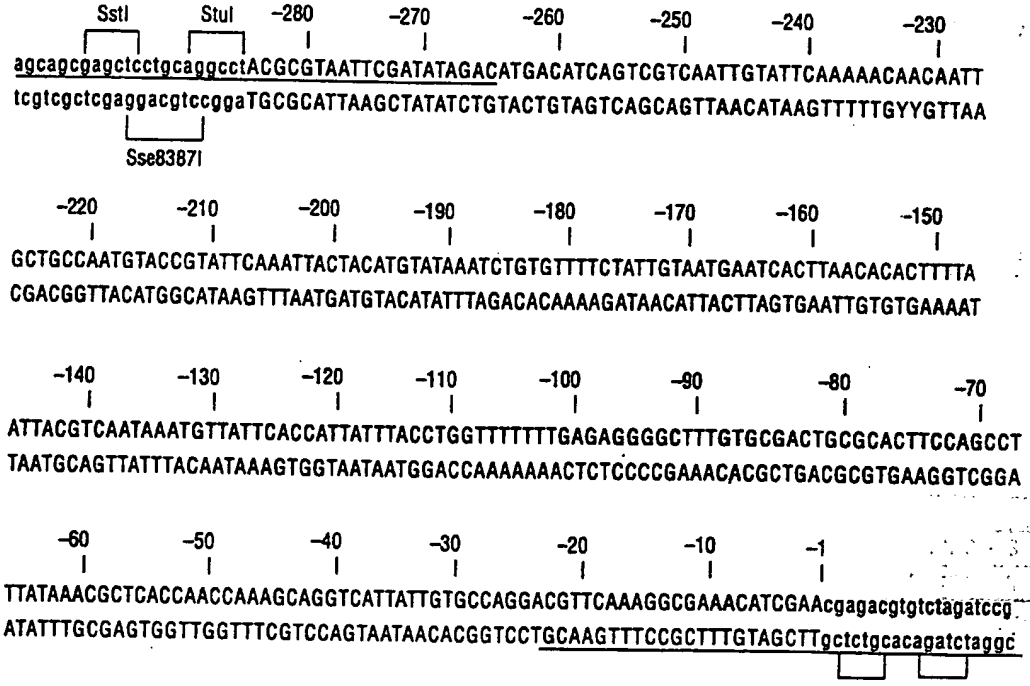


Figure 12

PCR AMPLIFICATION OF THE AcMNPV 35K GENE PROMOTER FOR Esp 3I-BASED MODULAR EXPRESSION VECTORS

PRIMER SET:

Strand	Primer	Sequence
(+) strand	35KPRO1	5' – AGCAGCGAGCTCCTGCAGGCCTCTTGATGTCTCCGATTTC – 3'
(–) strand	35KPRO2	5' – CGGATCTAGACACGTCTCGTTTGCTATGGTAAAGCTCAAA – 3'

TEMPLATE: AcMNPV Hind III fragment "K"

AMPLIFIED FRAGMENT:

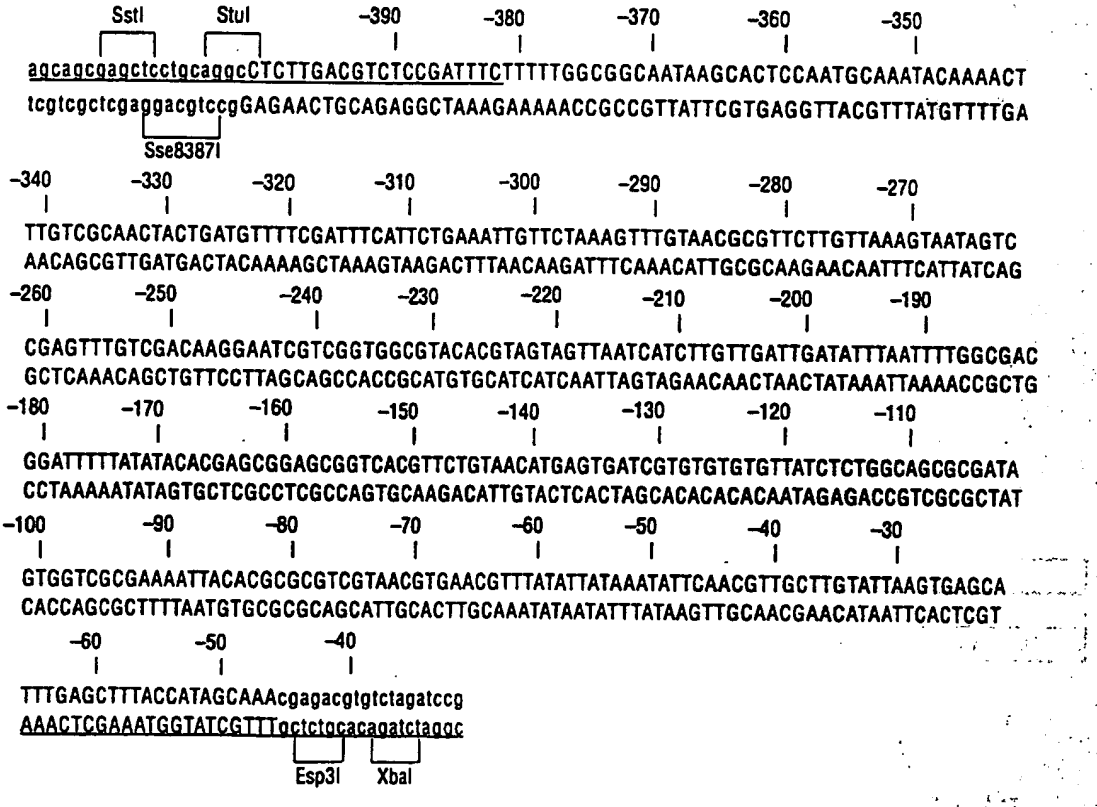


Figure 13

PCR AMPLIFICATION OF THE AcMNPV 6.9K GENE PROMOTER FOR Esp 3I-BASED MODULAR EXPRESSION VECTORS

PRIMER SET:

Strand	Primer	Sequence
(+) strand	69KFZ	5' – AGCAGCGAGCTCCTGCAGGCCTATGCCGTGTCCAATTGCAAG – 3'
(–) strand	69KRZ	5' – CGGATCTAGACACGTCTCGGTTTAAATTGTGTAATTTATGTA – 3'

TEMPLATE: AcMNPV Hind III fragment "H"

AMPLIFIED FRAGMENT:

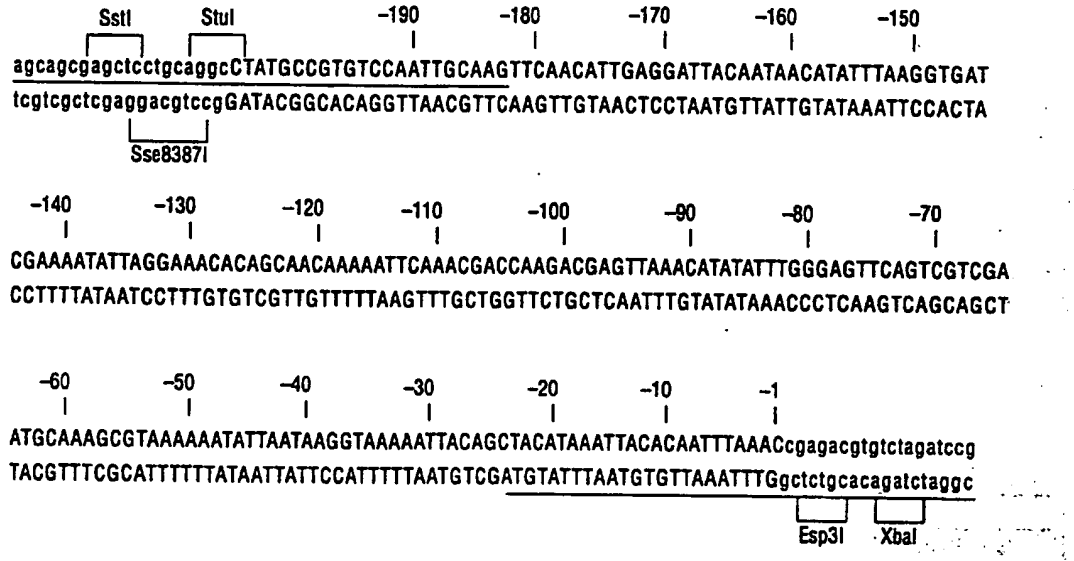


Figure 14

PCR AMPLIFICATION OF THE AcMNPV POLYHEDRIN GENE PROMOTER FOR Esp 3I-BASED MODULAR EXPRESSION VECTORS

PRIMER SET:

Strand	Primer	Sequence
(+) strand	PHF	5' – AGCAGCGAGCTCCTGCAGGCCTGACGCACAAACTAATATCAC – 3'
(–) strand	PHR	5' – CGGATCTAGACACGTCTCGATTTATAGGTTTTTTTATTACAA – 3'

TEMPLATE: AcMNPV Hind III fragment "F"

AMPLIFIED FRAGMENT:

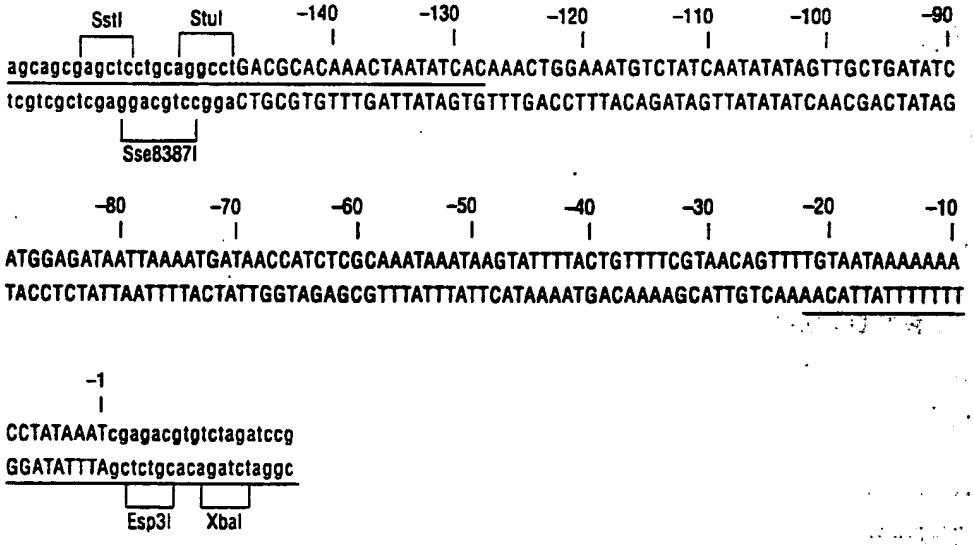


Figure 15
Sheet 1 of 2

PCR AMPLIFICATION OF THE DROSOPHILA hsp70 GENE PROMOTER FOR Esp 3I-BASED MODULAR EXPRESSION VECTORS

PRIMER SET:

Strand	Primer	Sequence
(+) strand	DMHSP70F	5' - AGCAGCGAGCTCCTGCAGGCCTGTCTTGGCTTTCCTCATATGTAC - 3'
(–) strand	DMHSP70R	5' - CGGATCTAGACACGTCTCGTGTGTGTGAGTTCTTCTTTCTCGG - 3'

TEMPLATE: *Drosophila melanogaster* locus 87C1

AMPLIFIED FRAGMENT:

```
          SstI      StuI      -800      -790      -780      -770      -760      -750
agcagcgagctcctgcaggcctGTCTTGGCTTTCCTCATATGTACATATGTATGTAAATATGTAAAATAAGTCGCAACTA
tcgtcgctcgaggacgtccggaCAGAACCGAAAGGAGTATACATGTATACATACATTTATACATTTTATTCAGCGTTGAT
                Sse8387I

          -740      -730      -720      -710      -700      -690      -680      -670
AATTCTAATACATTTTTCAGAATCTTAAATTAATTTTATCGTATATTAAAACAGAAGAAAGTCCGTTAATAGTTGATTTC
TTAAGATTATGTAAAAAGTCTTAGAATTTAATTAAAATAGCATATAATTTTGTCTTCTTTCAGGCAATTATCAACTAAAG

          -660      -650      -640      -630      -620      -610      -600      -590
ATTAACTAAAAGTACAAAATAATCTTTAATACATATGCCGATCAGACATTTATTGGTTTAGAAGCGCAGTATTTTTTTTG
TAATTGATTTTCATGTTTTATTAGAAATTATGTATACGGCTAGTCTGTAAATAACCAAATCTTCGCGTCATAAAAAAAAC

          -580      -570      -560      -550      -540      -530      -520      -510
CGAATACGCATAACAAAGCGCTTCGATTATCTTTAACATAAGTTATTTAAGCAGCCGTATTTATAAAGAAATTTCCAAAA
GCTTATGCGTATTGTTTCGCGAAGCTAATAGAAATTGTATTCAATAAATTCGTCGGCATAAATATTTCTTTAAAGGTTTT
```

Figure 15
Sheet 2 of 2

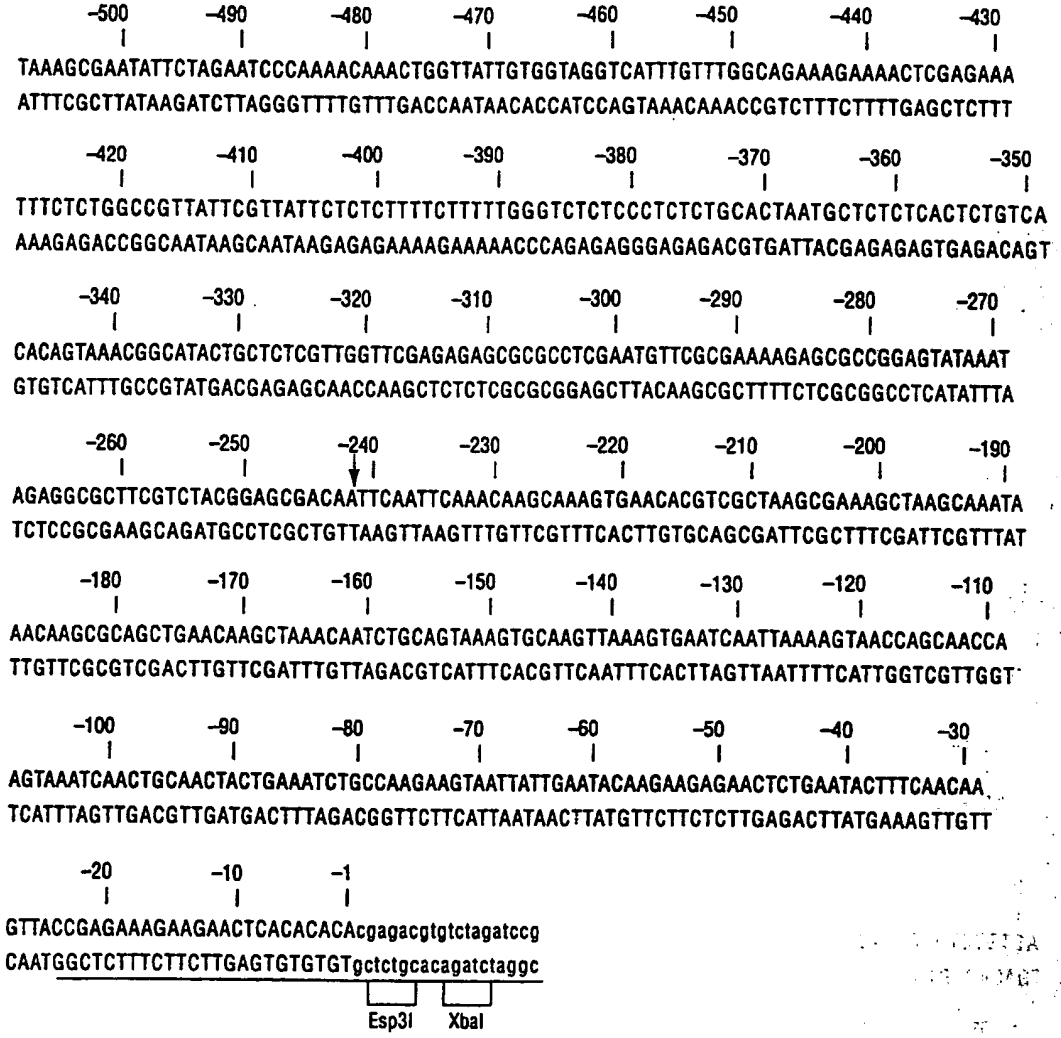


Figure 16
Sheet 1 of 2

PCR AMPLIFICATION OF THE AcMNPV 35K GENE 3'-UTR AND hr5 FOR Esp 3I-BASED MODULAR EXPRESSION VECTORS

PRIMER SET:

Strand	Primer	Sequence
(+) strand	35KHR5A	5' – ATTCCATGGATCCCGGGTACCTGTAACTAGTGCACTCAAC – 3'
(–) strand	35KHR5B	5' – AGCAGCCTCGAGCCTCAGGCCTCCACATTGTCGACTTGCT – 3'

TEMPLATE: AcMNPV Hind III fragment "Q"

AMPLIFIED FRAGMENT:

```
  NcoI   SmaI                                              35K gene end
                                                    35 K TER
attccatggatcccgggtaccTGTAACTAGTGCACTCAACAAAAATGTAATATTAAACACAATTAAATAAATGTTGAAA
taaggtacctagggcccatggACATTGATCACGTGAGTTGTTTTTACATTATAATTTGTGTTAATTTATTTACAACTTT
      BamHI  KpnI

ATTTATTGCCTAATATTATTTTTGTCAGTTSSTTGTCATTTATTAATTTGGATGATGTCCATTTGTTTTTAAAATTGAAC
TAAATAACGGATTATAATAAAAACAGTCAASSAACAGTAAATAATTAAACCTACTACAGGTAAACAAAAATTTTAACTTG

            EcoRI
TGGCTTTACGAGTAGAATTCTACGCGTAAAACACAATCAAGTAYGAGTCATAAGCTGATGTCATGTTTTGCACACGGCTC
ACCGAAATGCTCATCTTAAGATGCGCATTTTGTGTTAGTTCATRCTCAGTATTCGACTACAGTACAAAACGTGTGCCGAG

                      EcoRI
ATAACCGAACTGGCTTTACGAGTAGAATTCTACTTGTAACGCACGATCGAGTGGATGATGGTCATTTGTTTTTCAAATCG
TATTGGCTTGACCGAAATGCTCATCTTAAGATGAACATTGCGTGCTAGCTCACCTACTACCAGTAAACAAAAAGTTTAGC

                                                EcoRI
AGATGATGTCATGTTTTGCACACGGGCTCATAAACTCGCTTTACGAGTAGAATTCTACGTGTAACGCACGATCGATTGAT
TCTACTACAGTACAAAACGTGTGCCCGAGTATTTGAGCGAAATGCTCATCTTAAGATGCACATTGCGTGCTAGCTAACTA
```

Figure 16
Sheet 2 of 2

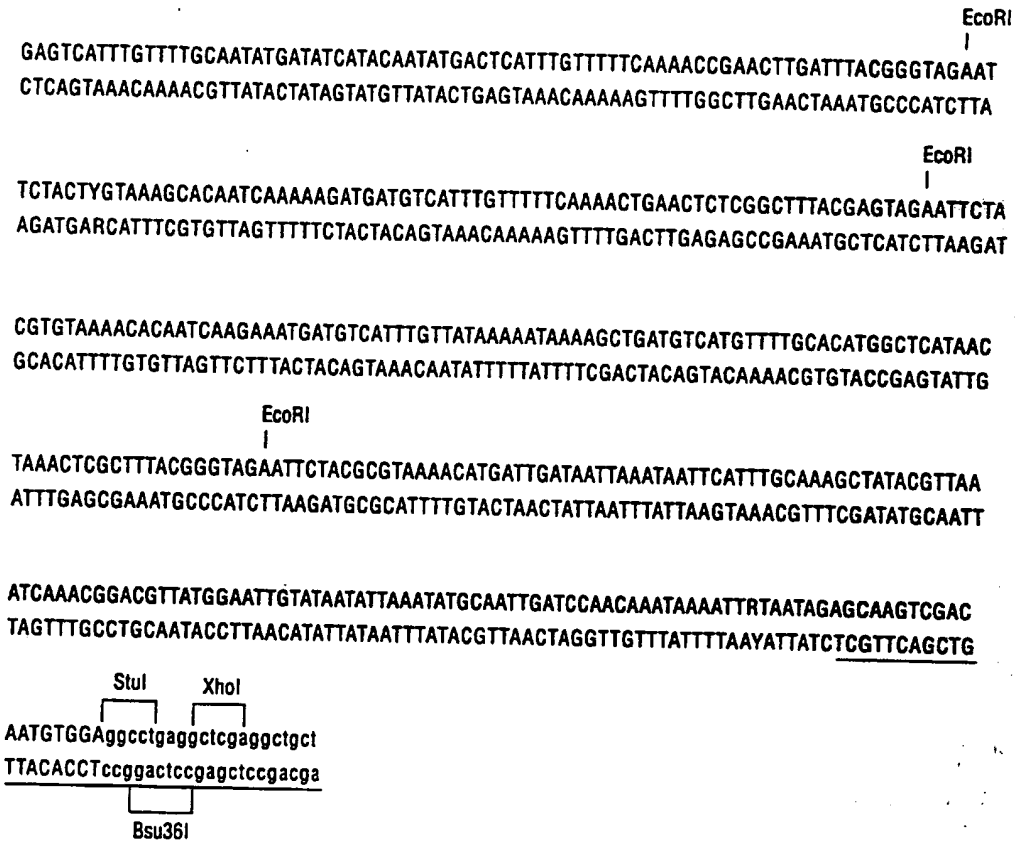


Figure 17A

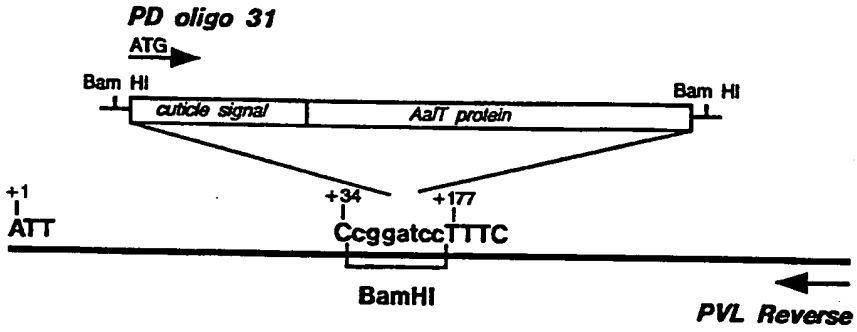

PD oligo 31
ATG
Bam HI
Bam HI
cuticle signal
AaIT protein
+1
ATT
+34
+177
CcggatccTTTC
BamHI
PVL Reverse

Figure 17B

PCR AMPLIFICATION OF THE CODON-OPTIMIZED CUTICLE-AalT GENE SEQUENCE FROM PLASMID AC0055.1 (transfer vector pVL985/Cuticle-AalT-1)

PRIMER SET:

Strand	Primer	Sequence
(+) strand	PD oligo 31	5' - ATGTTCAAGTTCGTGATGATCTGCGCC - 3'
(−) strand	PVLReverse	5' - GGATTTCCTTGAAGAGAGTGAG - 3'

TEMPLATE: Plasmid pAC0055.1 (pVL985/Cuticle-AalT-1)

AMPLIFIED PRODUCT:

Met Phe Lys Phe Val Met Ile Cys Ala Val Leu Gly Leu Ala Val Ala Lys Lys Asn Gly
ATG TTC AAG TTC GTG ATG ATC TGC GCC GTC CTC GGC CTG GCT GTG GCC AAG AAG AAC GGC
TAC AAG TTC AAG CAC TAC TAG ACG CGG CAG GAG CCG GAC CGA CAC CGG TTC TTC TTG CCG

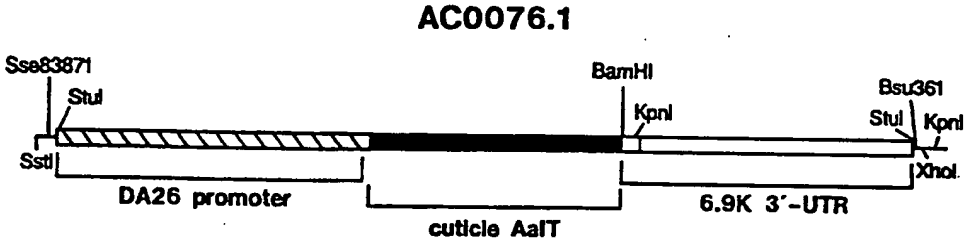
Tyr Ala Val Asp Ser Ser Gly Lys Ala Pro Glu Cys Leu Leu Ser Asn Tyr Cys Asn Asn
TAC GCA GTC GAC TCA TCC GGA AAA GCC CCC GAG TGC CTG CTC TCG AAC TAT TGC AAC AAT
ATG CGT CAG CTG AGT AGG CCT TTT CGG GGG CTC ACG GAC GAG AGC TTG ATA ACG TTG TTA

Glu Cys Thr Lys Val His Tyr Ala Asp Lys Gly Tyr Cys Cys Leu Leu Ser Cys Tyr Cys
GAA TGC ACC AAG GTG CAC TAC GCT GAC AAG GGC TAC TGT TGC CTT CTG TCC TGC TAT TGC
CTT ACG TGG TTC CAC GTG ATG CGA CTG TTC CCG ATG ACA ACG GAA GAC AGG ACG ATA ACG

Phe Gly Leu Asn Asp Asp Lys Lys Val Leu Glu Ile Ser Asp Thr Arg Lys Ser Tyr Cys
TTC GGT CTC AAC GAC GAC AAG AAA GTT CTG GAA ATC TCT GAT ACT CGC AAG AGC TAC TGT
AAG CCA GAG TTG CTG CTG TTC TTT CAA GAC CTT TAG AGA CTA TGA GCG TTC TCG ATG ACA

BamHI
Asp Thr Thr Ile Ile Asn TER ⌐
GAC ACC ACC ATC ATT AAC TAA GGATCCTTTCCTGGGACCCGGCAAGAACCAAAAACTCACTCTCTTCAAGGAAATCC
CTG TGG TGG TAG TAA TTG ATT CCTAGGAAAGGACCCTGGGCCGTTCTTGGTTTTTGAGTGAGAGAAGTTCCTTTAGG

Figure 18



SstI Sse36I StuI
GAGCTCCTGCAGGCCTACGCGTAATTCGATATAGACATGACATCAGTCGTCAATTGTATTCATTAAAAACAACAG
CTGCCAATGTACCGTATTCAAATTACTACATGTATAAATCTGTGTTTTCTATTGTAATGAATCACTTAACACACT
TTTAATTACGTCAATAAATGTTATTCACCATTATTTACCTGGTTTTTTTGAGAGGGGCTTTGTGCGACTGCGCAC
TTCCAGCCTTTATAAACGCTCACCAACCAAAGCAGGTCATTATTGTGCCAGGACGTTCAAAGGCGAAACATCGAA
ATG TTC AAG TTC GTG ATG ATC TGC GCC GTC CTC GGC CTG GCT GTG GCC AAG
AAG AAC GGC TAC GCA GTC GAC TCA TCC GGA AAA GCC CCC GAG TGC CTG CTC
TCG AAC TAT TGC AAC AAT GAA TGC ACC AAG GTG CAC TAC GCT GAC AAG GGC
TAC TGT TGC CTT CTG TCC TGC TAT TGC TTC GGT CTC AAC GAC GAC AAG AAA
GTT CTG GAA ATC TCT GAT ACT CGC AAG AGC TAC TGT GAC ACC ACC ATC ATT

BamHI KpnI
AAC TAA G GATCCCGGGTACCAACCAGACATTCCACACAGCCGACAGTAGCGAATGAACGAAGCGATTTCGT
CGCCTGCCCTCGTTTGGCTTTCGACTGTTACAAAATCATGTCTGCAAGATTTTAAACTAAGCCCGCTAAGCTCAA

StuI Bsu36I
ATAGTTTATTTTTATTACTGTTTTGTAAATAAATAACTTTATCATTCAATATTTGCCTGTGTTTGAGGCCTGAGG

XhoI KpnI
CTCGAGGGGGGGCCCGGTACC

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/06079

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C12N 7/00, 7/01, 15/00, 15/85; C07H 21/04
US CL :435/69.1, 235.1, 320.1; 536/24.2; 935/23, 32, 52, 57
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 235.1, 320.1; 536/24.2; 935/23, 32, 52, 57

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Dialog, Biosis, Biotech
Search terms: in vitro, ligation, vector, baculovirus

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	AU, A, 35291/89 (DEVAUCHELLE ET AL) 07 December 1989, see entire document.	1-16
Y	J. K. Setlow et al., "Genetic Engineering", Volume 8, published 1986 by Plenum Publishing Corporation, pages 277-298, see pages 284-286.	1-16

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 JULY 1994	05 AUG 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DAVID GUZO Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*